University of Alberta

Robust Integer-Valued Designs for Generalized Linear Models

by

Adeniyi Johnson Adewale ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of *Doctor of Philosophy*

in

Statistics

Department of *Mathematical and Statistical Sciences*

Edmonton, Alberta

Fall 2006

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-22980-4
Our file Notre référence
ISBN: 978-0-494-22980-4

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

ABSTRACT

We investigate the construction of designs for misspecified generalized linear models. Possible misspecification of generalized linear models includes linear predictor misspecification, use of an incorrect link function and an inadequate variance function. In this work we assume a finite design space and our interest is in the construction of integer-valued designs over the finite design space. The advantage of our integer-valued construction over the approximate design approach is that our designs are exact and thus readily implementable.

We begin with the problem of designing for models with a misspecified linear predictor. We adopt the average mean squared error of predictions over the design design space as the loss function. The complicated dependence of the loss function on the unknown contamination function renders the problem of designing for generalized linear models with misspecified linear predictor not easily amenable to the minimax treatment which has been successful in the context of linear models. We propose a new criterion for robust designs termed "minave" designs. Using the average (over the design space) mean squared error of predictions as the loss function, the minave design is the design that minimizes the average mean squared error of predictions over the specified contamination neighbourhood. This averaging was carried out using a procedure based on a singular value decomposition of the design matrix.

We give a holistic treatment to the problem of designing for misspecified generalized linear models by using the same approach for constructing designs when the link function is possibly misspecified and when there is overdispersion. The general approach is to derive the mean squared error of predictions-based criterion in each case of model misspecification. Having established the design criterion based on relevant statistical consideration we obtain the integer-valued designs such that the criterion is minimized. In all cases the problem is a nonlinear integer optimization problem. We employ the simulated annealing algorithm to solve the resulting optimization problem.

ACKNOWLEDGEMENTS

I deeply appreciate the support and encouragement of my supervisor Dr. D.P. Wiens. He introduced me to Robust Statistics and in my early days of graduate studies he helped augment my foundations in Statistical Theory. This dissertation has benefited from his insights and intellectual acumen. I am privileged to have worked under his supervision. I am very grateful for the financial support he extended to me through his research funds.

My appreciation also goes to Drs. Ivan Mizera, K.C. Carriere, Bryon Schmuland, Norman Beaulieu for serving on my dissertation examination committee and to the external examiner Dr. W.K. Wong. I am thankful for their careful review of my thesis and for the valuable comments they provided.

My profound thankfulness and love to my wife and friend, Lola Adewale, for her companionship, understanding and encouragement while I was working to complete this dissertation. I acknowledge her sacrifice of love at those "present-in-body-but-absent-in-mind" times. I am grateful for the support in various forms from family and friends, to mention a few: My parents - Mr. & Mrs. Adebimpe Adewale, my uncle and aunty - Mr. & Mrs. J.A. Ajeleye, Deacon & Mrs. J.O. Bolaji, Deacon & Mrs. E.O. Bolaji, Wisdom Agorde and Shane & Dena Harris.

This dissertation work and conference travels were funded through a postgraduate scholarship from the Natural Science and Engineering Research Council of Canada,

support from Dr. D.P. Wiens, assistantship and scholarship from the Department of Mathematical and Statistical Sciences and award from the Faculty of Graduate Studies and Research.

Above all, I give all glory, honour and praise to the Almighty God to whose faithfulness and unfailing mercy I owe the completion of this program.

To my parents
Mr. Adebimpe Adewale and Mrs. Ruth Ajekemi Adewale
for their love, support and unwavering commitment to my
educational pursuits.

TABLE OF CONTENTS

ABSTRACT . . . .

ACKNOWLEDGEMENTS . . . .

I INTRODUCTION 1

1.1 Generalized Linear Models 1

1.2 Classical Linear Regression Designs 4

1.3 Designs for Nonlinear Regression Models and Generalized Linear Models 7

1.4 Model Robust Designs 11

1.5 Dissertation Summary 18

1.6 Highlights of Contributions 23

II NEW CRITERIA FOR ROBUST INTEGER-VALUED DESIGNS IN LINEAR MODELS 30

2.1 Statistical model 31

2.2 Loss functions 32

2.2.1 "Minave" Mean Squared Error model-robust design criteria . 33

2.2.2 Bias-Constrained D-optimal model-robust design criteria . . 35

2.3 Numerical algorithms 38

2.3.1 Simulated annealing algorithm for Minave designs 38

2.3.2 Simulated annealing algorithm for Bias-Constrained D-optimal designs 42

2.4 Examples: Polynomial regression 44

2.5 Designs for multiple regression 48

2.6 Concluding remarks 53

III ROBUST DESIGNS FOR MISSPECIFIED LOGISTIC MODELS 56

3.1 Introduction 56

3.2 Loss Functions: Estimated and Averaged Mean Squared Errors of Prediction 62

3.3 Designs - Algorithm and Examples 66

3.3.1 Simulated Annealing 66

3.3.2 Examples 68

3.4 Case Study: Beetle Mortality Data 77

3.5 Conclusions 79

3.6 Appendix: Derivations 81

IV ROBUST DESIGNS FOR BINOMIAL MODELS WITH OVERDISPERSION AND MISSPECIFIED LINK FUNCTIONS 89
  4.1 Introduction 89
  4.2 Statistical Models 95
    4.2.1 Overdispersion Model 97
    4.2.2 True Link Model 98
  4.3 Loss Function and Algorithm 101
    4.3.1 Design Criterion 101
    4.3.2 Algorithm: Modified Simulated Annealing 103
  4.4 Examples 107
  4.5 Application: Calandra Granaria Data 110
  4.6 Concluding Remarks 111

V CONCLUSIONS AND APPLICATIONS TO POISSON DATA . 120
  5.1 Model-dependency of Optimal Designs 120
  5.2 Applications in Poisson Models 123
    5.2.1 Poisson Models to be Fitted 123

5.2.2 Departures from the Assumptions and Design Criteria for Protection Against Departures 124

5.2.3 Summary of Implications of Model Misspecifications 128

5.3 Computational Issues 130

5.4 Application to Ames Salmonella Assay 132

5.5 Choosing Design Parameters using Efficiency 134

LIST OF TABLES

1 Minave Designs for Partial Second-order Model. 50

2 Minave Designs for Full Second-order Model. 50

3 Comparing Restricted Designs with Unrestricted Design; $\rho = 0$. . . . 70

4 Designs for simple logistic model when the true model has a quadratic term. 72

5 Experimental design and response values. 73

6 Design for unknown contamination with $\beta_0 = 1$, $\beta_1 = 3$. 76

7 Beetle mortality data. 77

8 Toxicity of Ethylene Oxide to the Calandra Granaria. 92

9 Bliss's Beetle mortality data. 92

10 Comparing Designs for the Toxicity of Ethylene Oxide on Calandra Granaria. 111

11 Number of revertant colonies of salmonella (Y_i). 121

12 Optimal Designs Minimizing Average Variance of Predictions. 123

13 Designs for a Poisson Model with Misspecified Linear Predictor. . . . 134

14 Designs for a Poisson Model with Protection Against Overdispersion. 135

15 Designs for a Poisson Model with Misspecified Link Function. 135

LIST OF FIGURES

1 Simulated annealing trajectory for cubic regression design with $N = 40$, $n = 20$ and $\rho = 1$. 42

2 Integer-valued designs for cubic regression with $N = 40$ and $n = 20$ for (a) $\rho = 1$, (b) $\rho = .75$, (c) $\rho = .25$ and (d) $\rho = 0$. 43

3 Integer-valued designs for cubic regression with $N = 40$ and $n = 40$ for (a) $\rho = 1$, (b) $\rho = .75$, (c) $\rho = .25$ and (d) $\rho = 0$. 44

4 Integer-valued designs for cubic regression with $N = 40$ and $n = 60$ for (a) $\rho = 1$, (b) $\rho = .75$, (c) $\rho = .25$ and (d) $\rho = 0$. 45

5 Plot of efficiency loss (versus ρ) due to mere replications of the design points for $n = 20$ as opposed to an explicit construction of the designs for $n = 60$. 45

6 Robust designs for parameter estimation with (a) $\alpha = 0.0924$, (b) $\alpha = 0.5$ and prediction with (c) $\beta = 4.2068$ and (d) $\beta = 10$. 46

7 Minave designs for partial second-order model (a) $\rho = 1$, (b) $\rho = .75$, (c) $\rho = .5$, (d) $\rho = 0$. 49

8 Minave designs for full second-order model (a) $\rho = 1$, (b) $\rho = .75$, (c) $\rho = .5$, (d) $\rho = 0$. 49

9 Simulated Annealing Trajectory for Logistic Design with $\eta = 1 + 3x$, $x \in \mathcal{S} = [-1, 1]$, $\rho = 0$ and $(N = 40, n = 200)$ 68

10 Locally optimal designs minimizing $\mathcal{L}_{I,ave}$ when $\rho = 0$ (no contamination) with (a) $N = 20, n = 20$; (b) $N = 20, n = 200$; (c) $N = 40, n = 20$; (d) $N = 40, n = 200$. 69

11 Estimated contamination plot for Example 3. True (but unknown) form of contamination is quadratic. 73

12 Locally optimal designs in Example 4; (a) - (d): $(\beta_0, \beta_1) = (1, 1)$; (e)-(h): $(\beta_0, \beta_1) = (1, 3)$; (i)-(l): $(\beta_0, \beta_1) = (3, 1)$; (m)-(p): $(\beta_0, \beta_1) = (3, 3)$. 75

13 Robust Bayesian optimal design in Example 5 with $\rho = .25$ and parameters β_0 and β_1 having independent uniform priors over (a) $[.5, 1.5] \times [2.5, 3.5]$, (b) $[.5, 1.5] \times [1, 5]$, (c) $[-1, 3] \times [2.5, 3.5]$, (d) $[-1, 3] \times [1, 5]$. 77

14 (a) Prediction design when contamination is estimated from initial data. (b) Robust Bayesian prediction design with multivariate normal prior and $\rho = 5$. 79

15 Mortality of Calandra Granaria on exposure to ethylene oxide 93

16 Deviance versus λ for the beetle mortality data 93

17 Proportion of dead beetles versus dose with logit link and $\lambda = .1861$-link superimposed 94

18 Minave designs when (a) $\Phi = [0, .005]$, (b) $\Phi = [0, .05]$, (c) $\Phi = [0, .25]$, and (d) $\Phi = [0, .5]$ with $\Theta = [.5, 1.5] \times [2.5, 3.5]$ and $\Lambda = \{1\}$. 108

19 Minave designs when (a) $\Lambda = [.75, 1]$, (b) $\Lambda = [.25, 1]$, (c) $\Lambda = [0, 1]$, and (d) $\Lambda = [0, 2]$ with $\Theta = [.5, 1.5] \times [2.5, 3.5]$ and $\Phi = \{0\}$. 109

20 Number of revertant colonies of salmonella against dose level with Model I and II superimposed 121

21 Efficiency of robust designs for a Poisson model with misspecified linear predictor over a grid of values of ρ in the interval $[0, 100]$ 136

22 Efficiency of robust designs Poisson model with overdispersion 136

23 Efficiency of robust designs for a Poisson model with misspecified link function over a grid of values of λ in the interval $[0, 1]$ 137

CHAPTER I

INTRODUCTION

This dissertation is on the construction of optimal designs for generalized linear models when there is the possibility that the assumed form of the fitted model does not accurately represent the true model.

1.1 Generalized Linear Models

Generalized linear models are a body of regression-like models encompassing data from binomial, Poisson, gamma, and inverse Gaussian distributions in addition to the classical normal distribution. Generalized linear models are extensions of the ordinary regression models to non-normal responses (see McCullagh and Nelder, 1989). The response Y is assumed to have a distribution in the exponential family of distributions, taking a probability density/mass function of the form

$$f_Y(y;\theta,\phi) = \exp\left\{(y\theta - b(\theta))/a(\phi) + c(y,\phi)\right\} \tag{1.1}$$

for some functions $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$. The parameter θ is the canonical parameter. A generalized linear model has three components:

1. Random component: The random component of a generalized linear model specifies the distribution of the response variable Y. The distribution has the form (1.1) and for any distribution of this form the mean and variance of the

response variable Y are given by

$$E(Y) = \mu = b'(\theta)$$

and

$$var(Y) = a(\phi)\, b''(\theta),$$

where primes denote differentiation with respect to the canonical parameter θ. Thus the variance of the response is prescribed by the assumed distribution.

2. Systematic component: The systematic component defines a non-stochastic linear model

$$\eta = \mathbf{z}^T(\mathbf{x})\,\boldsymbol{\beta}$$

where $\mathbf{z}^T(\mathbf{x})$ is the regressor vector corresponding to the vector of explanatory variables $\mathbf{x}$ and $\boldsymbol{\beta}$ is the vector of model parameters. This linear combination of regressors is called the linear predictor in the generalized linear model literature.

3. Link component: The third component of a generalized linear model specifies a monotonic differentiable function $g(\cdot)$ termed the link function. The link function connects the random and systematic components. This connection is done by equating the mean response μ to the linear predictor η by $\eta = g(\mu)$, that is

$$g(\mu) = \mathbf{z}^T(\mathbf{x})\,\boldsymbol{\beta}.$$

The link function $g(\mu) = \mu$ is the identity link function which equates the mean response to the linear predictor. Thus, the link function for the ordinary regression with normally distributed response variable Y is the identity link.

The link function which equates the linear predictor to the canonical parameter is the canonical link. That is, $\eta = \mathbf{z}^T(\mathbf{x})\boldsymbol{\beta} = g(\mu) = \theta$.

Usually model fitting for generalized linear models is by maximum likelihood estimation. The likelihood equations are usually nonlinear in the model parameters $\boldsymbol{\beta}$. These equations are solved using iterative methods, in particular, Newton-Raphson method or Fisher scoring method. Details of the model fitting procedures can be found in the classical book by McCullagh and Nelder (1989) on generalized linear models.

This introductory chapter presents a literature review of regression design problems for linear and nonlinear models and the particular case of generalized linear models. Section 1.2 presents the classical regression design problem and the underlying assumptions. It includes a brief review of the various optimality criteria in the experimental design literature. Section 1.3 describes an extension of the regression design problem to nonlinear regression models and generalized linear models. It highlights the problem of parameter dependency of design criteria. This happens to be a major difficulty to be overcome when designing for nonlinear models and generalized linear models. Section 1.4 introduces the subject of model-robust designs. It includes a review of the pioneering work of Box and Draper (1959) in model-robust designs and the development of robust approaches sequel to Box and Draper's work. The section concludes with a case for the work done in this dissertation on the need for model-robust designs procedures in generalized linear models. Finally, we discuss the robustness issues around designs for generalized linear models and a summary of the chapters in this dissertation addressing these robustness issues.

1.2 Classical Linear Regression Designs

Regression is a statistical tool used for obtaining information on a response variable Y that depends on a (possibly vector valued) variable $\mathbf{x}$. When the variable $\mathbf{x}$ is under the control of an experimenter, he may like to know the values of $\mathbf{x}$ where it is "best" to observe the response Y. Usually, the experimenter is constrained by resources such as money, time and the number of observations he can take. The optimal regression design problem is about choosing levels of $\mathbf{x}$ and allocating observations at $\mathbf{x}$ so as to optimize specified criteria related to various constraints. There is a vast number of criteria in the experimental design literature. The choice of criteria would depend on the object of experimentation.

In the linear regression setting, the experimenter observes a response Y_i, given a vector of predictors $\mathbf{x}_i$ for a subject i, $i = 1, 2, ..., n$, assuming a model of the form

$$Y_i = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta} + \epsilon_i, \tag{1.2}$$

where $\boldsymbol{\beta}$ is the vector of unknown parameters and ϵ_i are random errors such that they are uncorrelated and have constant variance σ^2. The least squares estimate of the vector of model parameters is given by

$$\hat{\boldsymbol{\beta}} = \mathbf{M}^{-1}\mathbf{ZY},$$

where $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)^T$, $\mathbf{Z} = (\mathbf{z}(\mathbf{x}_1), \mathbf{z}(\mathbf{x}_2), ..., \mathbf{z}(\mathbf{x}_n))^T$ and $\mathbf{M} = \mathbf{Z}^T\mathbf{Z}$, the information matrix. The information matrix depends on the design vector $\mathbf{x}$ through the matrix $\mathbf{Z}$. The matrix $\mathbf{Z}$ is called the design matrix.

Suppose an experimenter would like to conduct an experiment whose response Y satisfies (1.2). When the total number of observations to be taken is n, the object of the optimal regression designs is to choose optimal values of $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$, not necessarily distinct, from a design space $\mathcal{S}$ such that certain criteria are satisfied. An $n-$tuple of points $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ from the design space $\mathcal{S}$ is an exact design. Thus the exact design corresponds to a discrete probability measure ξ on $\mathcal{S}$ with masses which are integral multiples of n^{-1}. Denote the information matrix of $\boldsymbol{\beta}$ corresponding to the design ξ as $\mathbf{M}(\xi)$. Finding exact designs is an integer optimization problem - optimization in a discrete domain - which is, in general, analytically intractable. The intractability of the exact problem led to the development of Kiefer's "approximate theory." In the approximate theory the restriction of the design measures to discrete probability measures is removed and a design measure could be any probability distribution ξ in a suitable class Ξ of distributions. The information matrix $\mathbf{M}(\xi)$ corresponding to a probability distribution ξ can be written as,

$$\mathbf{M}(\xi) = \int_{\mathcal{S}} \mathbf{z}(\mathbf{x})\, \mathbf{z}^T(\mathbf{x})\; d\xi.$$

The matrix $\mathbf{M}(\xi)$ is nonnegative definite and is assumed here to be positive definite. Optimal designs are usually obtained by optimizing functions of the information matrix, $\mathbf{M}(\xi)$, say $\Psi(\mathbf{M}(\xi))$.

The most intensively studied design criterion is the D-optimality criterion (Silvey 1980) and it is the design that maximizes the determinant of the information matrix.

That is

$$\xi = \underset{\xi \in \Xi}{\operatorname{argmax}} \det \{\mathbf{M}(\xi)\}.$$

This design minimizes the determinant of the variance-covariance matrix of the estimates of the model parameters. Other criteria that have been studied in the literature include the G-optimality criterion, the design minimizing the maximum (over the design space) variance of the predicted response (Kiefer and Wolfowitz, 1960). That is,

$$\xi = \min_{\xi \in \Xi} \max_{\mathbf{x} \in \mathcal{S}} \left\{\mathbf{z}^T(\mathbf{x})\, \mathbf{M}^{-1}(\xi)\, \mathbf{z}(\mathbf{x})\right\}.$$

The Q-optimality criterion, also known as I-optimal criterion seeks the design minimizing the integrated (or average) variance of the estimated response over the design space;

$$\xi = \min_{\xi \in \Xi} \int_S \mathbf{z}^T(\mathbf{x})\, \mathbf{M}^{-1}(\xi)\, \mathbf{z}(\mathbf{x})\, d\mathbf{x}.$$

The A-optimality criterion seeks the design minimizing the trace of the variance-covariance matrix;

$$\xi = \underset{\xi \in \Xi}{\operatorname{argmin}}\, trace \left\{\mathbf{M}^{-1}(\xi)\right\}.$$

The E-optimality criterion seeks the design minimizing the maximum eigenvalue of the variance-covariance matrix of model estimates (Kiefer, 1974);

$$\xi = \underset{\xi \in \Xi}{\operatorname{argmin}}\, \lambda_{\max} \left\{\mathbf{M}^{-1}(\xi)\right\}.$$

The c-optimality criterion seeks the design minimizing the variance of a given linear combination of parameter estimates. For a fixed vector $\mathbf{a}$, the c-optimal design is given by

$$\xi = \underset{\xi \in \Xi}{\operatorname{argmin}} \left\{ \mathbf{a}^T \mathbf{M}^{-1}(\xi)\, \mathbf{a} \right\}.$$

Kiefer and Wolfowitz (1960) present extensive results on D- and G-optimality, including the celebrated Equivalence Theorem. The Equivalence Theorem established that a design is D-optimal if and only if it is G-optimal. This means that the design that maximizes $\det\{\mathbf{M}(\xi)\}$ minimizes the maximum value of $\mathbf{z}^T(\mathbf{x})\,\mathbf{M}^{-1}(\xi)\,\mathbf{z}(\mathbf{x})$ over the design space $\mathcal{S}$. With approximate theory comes mathematical convenience such that the various optimizations which are otherwise unwieldy in the exact theory become tractable through convex theory. However, the resulting designs from approximate theory are not directly implementable. They need to be approximated by exact designs. The books by Fedorov (1972), Silvey (1980) and Pukelsheim (1993) are classical references on this subject.

1.3 Designs for Nonlinear Regression Models and Generalized Linear Models

A nonlinear regression model has a mean response of the form $E(y) = h(\mathbf{x}, \boldsymbol{\beta})$, with the function h a nonlinear function in at least one of the p model parameters. A constant variance is usually assumed. The information matrix for the vector of model parameters, $\boldsymbol{\beta}$ is given by

$$\mathbf{M}(\mathbf{x}, \boldsymbol{\beta}) = \sum \mathbf{z}(\mathbf{x}_i, \boldsymbol{\beta})\, \mathbf{z}^T(\mathbf{x}_i, \boldsymbol{\beta}) = \mathbf{Z}^T(\boldsymbol{\beta})\, \mathbf{Z}(\boldsymbol{\beta}),$$

where the vector,

$$\mathbf{z}(\mathbf{x}_i, \boldsymbol{\beta}) = \partial h(\mathbf{x}_i, \boldsymbol{\beta}) / \partial \boldsymbol{\beta}$$

and

$$\mathbf{Z}(\boldsymbol{\beta}) = \left(\mathbf{z}(\mathbf{x}_1, \boldsymbol{\beta}), \mathbf{z}(\mathbf{x}_2, \boldsymbol{\beta}), ..., \mathbf{z}(\mathbf{x}_n, \boldsymbol{\beta})\right)^T.$$

In the context of the approximate theory, the information matrix is given by

$$\mathbf{M}(\xi, \boldsymbol{\beta}) = \int_S \mathbf{z}(\mathbf{x}, \boldsymbol{\beta})\, \mathbf{z}^T(\mathbf{x}, \boldsymbol{\beta})\; d\boldsymbol{\xi},$$

for a continuous design $\boldsymbol{\xi}$. As it is with linear models, designing for a nonlinear model requires optimizing a certain function, say $\Psi(\mathbf{M}(\xi, \boldsymbol{\beta}))$, of the information matrix.

Nonlinear experimental design poses a major challenge because of the dependency of design criteria on the unknown parameter, $\boldsymbol{\beta}$, through the information matrix. Thus, designing an experiment for the estimation of the model parameters, $\boldsymbol{\beta}$, requires that these parameters be known! Various strategies have been proposed and used for dealing with the dependency of the design criteria on model parameters. The easiest and earliest approach is to adopt a best guess of the parameter values, say $\boldsymbol{\beta}^o$. Given best guesses for parameter values, the nonlinear design problem becomes amenable to the theory of optimal design for linear models. Chernoff (1953) dubbed this design locally optimal design. Locally optimal design can be very sensitive to the choice of the best guesses for the parameters. An approach that has been used to remedy the non-robustness of the locally optimal design is a Bayesian paradigm. In the Bayesian approach a prior distribution, say $\pi(\boldsymbol{\beta})$, is assumed on the unknown parameters. The Bayesian optimal design is the design optimizing the expectation of the criterion of interest, where expectation is taken with respect to the assumed prior distribution. That is,

$$E_{\beta}\Psi(\mathbf{M}(\xi, \boldsymbol{\beta})) = \int \Psi(\mathbf{M}(\xi, \boldsymbol{\beta}))\, \pi(\boldsymbol{\beta})\; d\boldsymbol{\beta}.$$

The prior distribution is usually interpreted as the experimenter's prior belief in the adequacy of the model over a specified range of parameter values. Chaloner and Larntz (1989), Chaloner and Verdinelli (1995) and Dette and Wong (1996), among others, have studied Bayesian designs.

An alternative to the Bayesian paradigm is the minimax (or maximin) approach used by Sitter (1992). The approach assumed that there is range of plausible values for unknown parameters. That is, $\boldsymbol{\beta} \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is a range of specific (not represented by distribution) parameter values the experimenter beliefs are plausible. The minimax optimal design is the design minimizing the maximum (over the range of the parameters) of the criterion, that is,

$$\min_{\xi} \max_{\boldsymbol{\beta} \in \boldsymbol{\Theta}} \Psi\left(\mathbf{M}\left(\xi, \boldsymbol{\beta}\right)\right).$$

This approach is robust in the sense that it produces the design corresponding to the worst possible parameter values within the range $\boldsymbol{\Theta}$. King and Wong (2000), Dette, Haines and Imhof (2003) and Biedermann, Dette and Pepelyvshev (2004) have also studied the minimax (or maximin) approach to address parameter-dependency of design criteria.

Sequential design is another strategy that has been used in dealing with parameter-dependency of design criteria. In sequential design, the experiment is done in stages. Parameter estimates from a previous stage are used as best guesses for the current design. Sequential design can be described as progressive locally optimal design. Abdelbasit and Plackett (1983) and Sinha and Wiens (2002) are among authors that have taken this approach to nonlinear design.

This thesis is focussed on designing for generalized linear models, a body of regression-like models with the response assuming a distribution from the exponential family of distributions (McCullagh and Nelder, 1989). Models from this class of models are mostly, but not exclusively, nonlinear models. This is so in the sense that the mean response, $E(Y) = \mu$, is usually a nonlinear function of the model parameters through a linear predictor, $\eta = \mathbf{z}(\mathbf{x})\boldsymbol{\beta}$. The variance of the response, $var(Y|x)$, is also a function of the linear predictor. The information matrix of $\boldsymbol{\beta}$ from a design comprising the points, $\mathbf{x}_i$, $i = 1, ..., n$, is given by

$$\sum_{i=1}^{n} w(\mathbf{x}_{i,}\ \boldsymbol{\beta})\mathbf{z}(\mathbf{x}_i)\,\mathbf{z}^T(\mathbf{x}_i) = \mathbf{Z}^T\mathbf{W}\mathbf{Z}$$

where

$$\mathbf{Z} = (\mathbf{z}(\mathbf{x}_1), \mathbf{z}(\mathbf{x}_2), ..., \mathbf{z}(\mathbf{x}_n))^T,$$

$$w(\mathbf{x}_{i,}\boldsymbol{\beta}) = \frac{1}{var(Y|\mathbf{x}_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2,$$

and

$$\mathbf{W} = diag(w(\mathbf{x}_{1,}\boldsymbol{\beta}), w(\mathbf{x}_{2,}\boldsymbol{\beta}), ..., w(\mathbf{x}_{n,}\boldsymbol{\beta})).$$

Thus, as with nonlinear experiments the information matrix depends on the unknown parameters, $\boldsymbol{\beta}$, through $w(\mathbf{x}_{i,}\ \boldsymbol{\beta})$ and the same treatment given to nonlinear regression models is applicable.

Of the nonlinear models belonging to this class of generalized linear models, after nonlinear model with normally distributed error, the binary models have received more attention in the experimental design literature. Abdelbasit and Plackett (1983), Khan and Yazdi (1988), Minkin (1987), Biedermann, Dette and Pepelyshev (2004), among others, have considered designs for binary models. Nevertheless, a handful of others have investigated designs for Poisson regression model (Minkin, 1993) and designs for the inverse Gaussian regression model (Fries and Bhattacharyya, 1986).

1.4 Model Robust Designs

As described in the previous sections, the objective in regression design is to take observations so as to fit a particular model. However, in most applications, the assumed model is, at best, only a reasonable approximation to the true model. The work of Box and Draper (1959) was the first to highlight, in a systematic way, the inherent dangers of designing a regression experiment on the basis that the assumed model is exactly correct. The disturbing feature of classical-regression designs is that they are dependent on the assumed model such that the optimal designs provide no opportunity to check the adequacy of the assumed model. Box and Draper studied the case where the experimenter fits a simple linear regression model whereas the true response has a quadratic term included. They assume that the true mean response is

$$\begin{aligned} E\left[Y\left(x\right)\right] &= \beta_0 + \beta_1 x + \beta_2 x^2, \\ \text{with } x &\in \mathcal{S} = [-1, 1] \end{aligned}$$

but the simple linear model,

$$\hat{Y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x$$

is fitted instead. Box and Draper used the integrated mean squared error of the fitted response as the design criterion. The integrated mean squared error is given by

$$\begin{aligned} I &= \frac{1}{\sigma^2}\int_S E\left[\hat{Y}(x) - E\left[Y(x)\right]\right]^2 dx \\ &= \frac{1}{\sigma^2}\int_S var\left[\hat{Y}(x)\right] + \frac{1}{\sigma^2}\int_S \left\{E\left[\hat{Y}(x)\right] - E\left[Y(x)\right]\right\}^2 dx \\ &= V + B, \end{aligned}$$

where V is the variance error due to sampling variation and B the squared bias error engendered by the model misspecification. Letting m_i denote the ith moment of the design, V and B can be expressed as follows:

$$\begin{aligned} V &= 1 + \frac{1}{3m_2}, \\ B &= \frac{N\beta_2^2}{\sigma^2}\left[m_2^2 - \frac{2m_2}{3} + \frac{1}{5} + \frac{m_3^2}{3m_2^2}\right]. \end{aligned}$$

The design minimizing I is the design with $m_3 = 0$ regardless of the values of m_2 and $\frac{N\beta_2^2}{\sigma^2}$. On setting $m_3 = 0$, I can be written as

$$I = \left[1 + \frac{1}{3m_2}\right] + \frac{N\beta_2^2}{\sigma^2}\left[\left(m_2 - \frac{1}{3}\right)^2 - \frac{4}{45}\right].$$

The optimal design then depends on $\frac{\beta_2}{\sigma/\sqrt{N}}$, a term Box and Draper called “a measure of the ratio of the quadratic curvature to the sampling error”. When this ratio is small, the true model is essentially a straight line up to sampling error and the corresponding design is about the classical optimal design which puts all its mass on -1 and +1. When this ratio is very large, it connotes that the straight line is a very poor approximation to the true model and the corresponding design is the design minimizing the bias B alone; this is what Box and Draper termed "all-bias design".

Box and Draper conclude that the designs minimizing the bias alone are closer to minimizing both the bias and variance than the designs minimizing variance alone.

Huber (1975) criticized the approach of Box and Draper (1959) that it only safeguards against arbitrarily selected polynomial alternatives but fails to safeguard against all potentially dangerous small deviations from the fitted model. Huber (1975) then proposed an alternative approach to robust design: suppose the true model is defined as

$$E\left[Y\left(x\right)\right] = \mathbf{z}^{T}\left(x\right)\boldsymbol{\beta} + f\left(x\right),$$

where the vector of regressor $\mathbf{z}\left(x\right) = \left(1,\ x\right)^{T}$, $x \in \mathcal{S} = \left[-1/2, 1/2\right]$ and the contamination function f belong to an infinite dimensional space of functions $\mathcal{F}$,

$$\mathcal{F} = \left\{ f : \int_{S} f^{2}\left(x\right)\,dx \leq \eta^{2},\ \int_{S} \mathbf{z}\left(x\right) f\left(x\right)\,dx = \mathbf{0} \right\}.$$

Thus, the true mean regression response is the simple linear regression function plus some unknown contamination function $f\left(x\right)$. Under this setting, the bias of the least squares estimate of model parameters is given by

$$\mathbf{bias} = E\left(\hat{\boldsymbol{\beta}}\right) - \boldsymbol{\beta} = \mathbf{B}^{-1}(\boldsymbol{\xi})\,\mathbf{b}\left(f, \xi\right),$$

where $\mathbf{b}\left(f,\xi\right) = \frac{1}{n}\sum \mathbf{z}\left(x_i\right) f\left(x_i\right) = \int_{S} \mathbf{z}\left(x\right) f\left(x\right) d\xi$. Thus, the mean squared error of the estimate $\hat{\boldsymbol{\beta}}$, is given by

$$\begin{aligned} MSE\left(f,\xi\right) &= E\left\{\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)^{T}\right\}, \\ &= cov\left(\hat{\boldsymbol{\beta}}\right) + \mathbf{bias}\cdot\mathbf{bias}^{T}, \\ &= \frac{\sigma^{2}}{n}\mathbf{M}^{-1}\left(\xi\right) + \mathbf{M}^{-1}\left(\xi\right)\mathbf{b}\left(f,\xi\right)\mathbf{b}^{T}\left(f,\xi\right)\mathbf{M}^{-1}\left(\xi\right). \end{aligned}$$

Huber used the integrated mean squared error

$$\text{IMSE} = \int_S E\left[\left(\mathbf{z}^T(x)\,\hat{\boldsymbol{\beta}} - E\left[Y(x)\right]\right)^2\right] dx,$$

as the design criterion. He obtained the design that minimizes the maximum integrated mean squared error over the class $\mathcal{F}$ of contamination functions, that is which solves

$$\min_{\xi\in\Xi} \max_{f\in\mathcal{F}} \text{IMSE}(\xi, f).$$

The resulting design has a density given by

$$m_0(x) = \left(ax^2 + b\right)^+, \tag{1.3}$$

where $(\cdot)^+ = \max(0,\cdot)$ and the constants a and b depend on the design parameter $v = (\sigma^2/n\eta^2)$. Huber's design, which he termed minimax design, is the design which is optimal for the worst possible contamination function $f \in \mathcal{F}$. Thus the design is reasonable for all $f \in \mathcal{F}$, meaning that it is a robust design. As $\nu \to \infty$, the resulting design approaches the classical Q-optimal design which places equal mass on the two extremes of $\mathcal{S}$ and as $\nu \to 0$ the resulting design approaches the uniform design. Wiens (1990) generalized Huber's work from simple linear regression to multiple linear regression where the vector of regressors is $\mathbf{z}(\mathbf{x}) = (1, x_1, ..., x_p)^T$. In a particular example on two interacting regressors with $\mathbf{z}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)^T$, $\mathcal{S} = [-1/2, 1/2] \times [-1/2, 1/2]$, Wiens (1990) obtained the optimal minimax design density as

$$m_0(x_1, x_2) = \left(a + b\left(x_1^2 + x_2^2\right) + cx_1^2x_2^2\right)^+,$$

where a, b and c depends on $v = (\sigma^2/n\eta^2)$.

The drawback of this approach is that it does not result in implementable designs but to "designs" that are rather arbitrary and possibly continuous probability functions with densities of the form $m_0(\mathbf{x})$. Hossain (2002) investigated strategies for implementing designs for simple linear regression from the design density (1.3). Heo, Schmuland and Wiens (2001) and Xu (2006) also discussed implementation strategies for multiple regression design problems.

Fang and Wiens (2000) while still using the Huber-type contamination class were able to obtain minimax designs that are integer-valued - implementable designs. Thus, they eliminated the drawback of the earlier minimax approach of Huber (1975) and Wiens (1990) which usually result in non-implementable designs. Fang and Wiens considered a finite design space $\mathcal{S}$ of points $\mathbf{x}_i$ $(i = 1, ..., N)$. So their design problem was to allocate non-negative integers n_i observations to $\mathbf{x}_i$ such that $\sum_{i=1}^{N} n_i = n$, the total number of observations. They use the average mean squared error, I of $\hat{Y}(x) = \mathbf{z}^T(x)\hat{\boldsymbol{\beta}}$ as the loss function

$$I = \frac{1}{N}\sum_{i=1}^{N} E\left[\left(\hat{Y}(\mathbf{x}_i) - E[Y|\mathbf{x}_i]\right)^2\right].$$

After relevant calculations, the loss function was expressed in the form

$$I = \frac{1}{N} tr\left(\mathbf{ZCZ}^T\right) + \frac{1}{N}\mathbf{f}^T\mathbf{Q}(\mathbf{Z},\mathbf{P})\,\mathbf{f} + \frac{1}{N}\mathbf{f}^T\mathbf{f}, \tag{1.4}$$

where $\mathbf{Z}$ is the matrix with rows $\mathbf{z}^T(\mathbf{x}_1), ..., \mathbf{z}^T(\mathbf{x}_N)$, $\mathbf{P}$ is a diagonal matrix with entries $\{p_i = n_i/n\}_{i=1}^{N}$, $\mathbf{Q}$ is a matrix which depends on $\mathbf{Z}$ and $\mathbf{P}$, $\mathbf{C} = cov\left[\hat{\boldsymbol{\beta}}\right]$ and $\mathbf{f} = (f(\mathbf{x}_1), ..., f(\mathbf{x}_N))$. The singular value decomposition of $\mathbf{Z}$, $\mathbf{Z} = \mathbf{U}_{N\times p}\mathbf{\Lambda}_{p\times p}\mathbf{V}^T_{p\times p}$ say, with $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}_p$ and $\mathbf{\Lambda}$ diagonal and invertible, is obtained. The $N \times p$

matrix $\mathbf{U}$ is augmented by $\tilde{\mathbf{U}}_{N\times(N-p)}$ such that $\left[\mathbf{U}\vdots\tilde{\mathbf{U}}\right]_{N\times N}$ is orthogonal. Then by the conditions defining the contamination class $\mathcal{F}$, they have that there is an $N-p\times 1$ vector $\mathbf{c}$, with $\|\mathbf{c}\|\leq 1$, satisfying $\mathbf{f}=\tau\sqrt{N}\tilde{\mathbf{U}}\mathbf{c}$. With this they obtained the maximization of (1.4) over f using simple matrix algebra. The results depend on η^2 and σ^2 through $v=\sigma^2/\left(n\eta^2\right)$ just like in the earlier approach which results in design densities. A simulated annealing algorithm is used to obtain the allocation of the non-negative integers n_i observations to $\mathbf{x}_i$ such that minimum of the maximum loss function is obtained. We adopt this approach to obtained integer-valued designs in this dissertation.

The criticism against the class of contamination functions used by Huber (1975) and Wiens (1990, 1992, 1998) is that the class is too broad because only designs which are absolutely continuous on $\mathcal{S}$ have a finite loss. Authors such as Marcus and Sacks (1976), Sacks and Ylvisaker (1978), Pesotchinsky (1982), Li and Notz (1982), Li (1984), and Liu and Wiens (1997) consider a different class of contamination functions:

$$\mathcal{F}=\left\{f:\left|f\left(\mathbf{x}\right)\right|\leq\phi\left(\mathbf{x}\right),\ \text{ for all }\mathbf{x}\in\mathcal{S}\right\},$$

with various assumptions about ϕ. The resulting designs from this class of contamination are generally supported on a small number of points and thus suffer the same fate has the classical optimal designs in that they often do not allow for test of lack of fit or the possibility of fitting more complex models.

Yue and Hickernell (1999) recently introduced a new contamination neighbourhood $\mathcal{F}$ which they assumed to be a Hilbert space of functions defined on the design space $\mathcal{S}$. Associated with the space is a symmetric function $K\left(\mathbf{x},\mathbf{w}\right)$, reproducing kernel, defined on $\mathcal{S} \times \mathcal{S}$, which has an inner product $\langle \cdot, \cdot \rangle_{\mathcal{F}_K}$. They also used the integrated mean squared error as the design criterion. Taking advantage of the properties of this reproducing kernel Hilbert space Yue and Hickernell place a sharp upper bound on the integrated mean squared error and then obtain the design minimizing this upper bound. Their approach to model-robust design protects against a broad class of alternatives like those using the Huber-type class of contamination and also allows the experimenter to obtain designs ranging from the all-variance design to the all-bias design by varying the design parameter. The approach overcomes the drawback of the Huber-type contamination in that it does not leave the experimenter with design densities but results in exact designs.

Given the preponderance of proposals and in fact substantial contributions to model-robust designs for linear models, it is surprising that very scant attention has been paid to model-robust designs for nonlinear models. It is not a far stretch to opine that the findings of Box and Draper (1959) in the linear models would be equally true for nonlinear models as well - designs constructed on the basis that the assumed model is exact portends an inherent danger. Ford, Titterington and Kitsos (1989) presented an expository work on designs for nonlinear models wherein they asserted that "indeed, if the model is seriously in doubt, the forms of design that we have considered may be completely inappropriate." Sinha and Wiens (2002) have explored some designs for nonlinear models when the assumed model is possibly misspecified. More recently, Wiens and Xu (2006) presented an article titled "Robust Prediction and Extrapolation Designs for Misspecified Generalized Linear Regression Models." They employed the minimax approach to robust designs with the notion that the true but unknown mean response is a nonlinear function (of model parameters) with an additive contamination function f which belongs to the Huber-type class of contamination functions discussed above. This thesis is on the construction of robust designs for generalized linear models with emphasis on binomial models although the methodologies presented are applicable to any generalized linear model.

1.5 Dissertation Summary

In this dissertation, we investigate the construction of robust designs for generalized linear models with due consideration for possible misspecification in the assumed model. From the experimenter's point of view, the design for generalized linear models involves the following procedure:

(a) choosing an appropriate error distribution,

(b) determining which variables to include in the systematic component, and

(c) defining the link function, $g(\mu)$.

Issues for robustness arise in any of these specifications. For the most part, the nature of the response does suggest which error distribution is reasonable. For instance, if the response is a proportion, the default error distribution would be the binomial distribution; for count data, the Poisson distribution would be the distribution of interest and for nonnegative-valued continuous data, the gamma or inverse Gaussian distributions would be of interest. Having specified the distribution, the form of the variance which is usually a function of the mean response, except for the normal distribution, is invariably specified. The need for robustness would arise when there is a chance that the data might exhibit more (or less) variability than that prescribed by the assumed distribution - a phenomenon called overdispersion (or underdispersion). Other than this, the specification of the distribution is usually not an issue from a design perspective. The systematic component is the linear predictor through which the mean response is connected to the design variables. The systematic component is subject to misspecification just as discussed in the context of linear models above. For example, the systematic component is misspecified when it does not reflect the influence of the design variables correctly. Usually there are many possibilities for the choice of a link function but the canonical link is commonly used - the identity link for normal models, logistic link for binomial models, the log link for Poisson models, and so on. An issue for robustness relating to the link function arises when the canonical link or any other choice is made when in fact a different link is more appropriate. For example, using the symmetric logistic link when the complementary log-log link is more appropriate.

In this work we present a holistic approach to designing for a generalized linear model when the specification of the model is possibly misspecified in any of the three components that defines the model. We start the pursuit of the objective of the dissertation - constructing robust designs for misspecified generalized linear models - with designs for logistic models with misspecified linear predictor. We seek for exact designs - implementable integer-valued designs - using the strategy proposed by Fang and Wiens (2000) within a finite design space, $\mathcal{S} = \{\mathbf{x}_i\}_{i=1}^{N}$, framework. The basis of the design criteria used is the mean squared error of predicted response. The problem invariably becomes an integer optimization problem for which we employ the simulated annealing algorithm. Given that the problem of robustness for the systematic component of a generalized linear model is akin to the many contributions on model-robust designs for linear models, we are comfortable starting out with this problem.

Suppose an experimenter fits a generalized linear model with the log-likelihood function

$$l_i(\boldsymbol{\beta}) = \sum \left\{ \left[y_i \theta_i - b(\theta_i) \right] / a(\phi) + c(y_i, \phi) \right\}, \ i = 1, ..., n,$$

with the mean response

$$\mu_i = b'(\theta_i) = \mu(\eta_i), \tag{1.5}$$

for $\eta_i = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta}$ when in fact the true mean response is represented by

$$\mu_{T,i} = \mu(\eta_i + f(\mathbf{x}_i)) \tag{1.6}$$

with

$$\mathcal{F} = \left\{ f : \frac{1}{N} \sum_{i=1}^{N} \mathbf{z}(\mathbf{x}_i) f(\mathbf{x}_i) = \mathbf{0}, \ \frac{1}{N} \sum_{i=1}^{N} f^2(\mathbf{x}_i) \leq \tau^2 \text{ with } \tau^2 = O\left(n^{-1}\right) \right\}. \tag{1.7}$$

It turns out that under this specification the asymptotic bias and variance of the maximum likelihood estimate of the model parameter $\boldsymbol{\beta}$ depends on the contamination function $f(\mathbf{x})$ through the true mean response $\mu(\eta + f(\mathbf{x}))$. Thus, the mean squared error of the estimate of model parameter $\boldsymbol{\beta}$ depends on the contamination function $f(\mathbf{x})$ in a complicated manner through both the bias term and the variance term. Hence, the maximization over the class of contamination functions, $\mathcal{F}$ required in the minimax approach to robust design seems hopeless in this context despite its successful implementation in linear models. See Huber (1975) and Wiens (1990, 1992, 2000). This problem motivated the development of new criteria for robust design.

In Chapter 2, we start by investigating these new criteria in the context of linear models. One of these new criteria is the "minave" approach to model-robust design. This approach is based on a notion of averaging of the loss function over the contamination neighbourhood $\mathcal{F}$. We found that this approach was very successful in constructing robust designs for linear models. The resulting designs have characteristics that are similar to the those of Wiens and others that have used the minimax approach. In addition to the "minave" approach we investigated constrained D-optimal design where the optimization is based on the classical D-optimal design criterion but subject to some bias-related constraints.

We present an article titled "Robust Designs for Misspecified Logistic Models" in Chapter 3. We consider the problem of designing when the linear predictor of the fitted logistic model is possibly misspecified. The design problem here is not readily amenable to the minimax approach to robust designs which has been successfully employed in linear models (Huber, 1975 and Wiens, 1990, 1992). The main challenge is the dependency of the variance and bias part of the design criterion on the unknown contamination function in a nonlinear manner. However, the minave tool developed for linear models in Chapter 2 is successfully adapted to proffer solution for designing for logistic models with misspecified linear predictors. Examples of robust designs for logistic models are presented, including a case-study implementing the methodologies using beetle mortality data.

Chapter 4 deals with other kinds of misspecifications in binomial models. In particular, we propose approaches for designing for binomial models when there is possible link function misspecification and overdispersion. Our framework assumes that the experimenter fits a canonical model when in fact the true model could be different, either because it has a different link function other than the canonical link function or because the variance prescribed by the assumed generalized linear model does not accurately describe the variability in the anticipated data. The logistic model, which corresponds to the canonical link for the binomial distribution, is often used for modelling binomial data for various reasons such as simplicity of interpretation in terms of odds-ratios. We note that misspecified link function can engender biased prediction in binomial models and propose a robust approach to design which contemplates that the true unknown link function belongs to a generalized family of links. Using an approach akin to that proposed by Pregibon (1980) on goodness of link tests for generalized linear models, the problem of designing when there is link function misspecification is recast as a linear predictor misspecification problem similar to that considered in Chapter 3. In order to construct design for binomial models with overdispersion, we extend the linear predictor of the true model to have a random contamination. That is,

$$\eta_i = \mathbf{z}^T(\mathbf{x})\,\boldsymbol{\beta}_0 + \phi v,$$

where $E(v) = 0$ and $var(v) = 1$. The true model is now somewhat a mixed-model in the spirit of the articles of Pierce and Sand (1975) and Williams (1982) on modelling for extra-binomial variation.

Chapter 5 concludes the dissertation work by putting the results of Chapters 3 and 4 in the framework of generalized linear models. In particular we consider applications in Poisson data.

Chapters 2, 3 and 4 are three independent articles which have been prepared for publication.

1.6 Highlights of Contributions

This thesis extends the concept of robustness in regression designs to designs for generalized linear models. It addresses three key robustness issues on misspecification of generalized linear models. The first of these deals with constructing designs for generalized linear models with a possibly misspecified linear predictor. The true linear predictor is assumed to differ from the fitted linear predictor by an additive contamination function belonging to the contamination neighbourhood defined by (1.7). As a loss function, the average mean squared error of predicted response is employed. This corresponds to the classical notion of I-optimality ("integrated variance"), which is also known as Q-optimality (Fedorov, 1972). The dependence of the average mean squared error of the predicted response on the contamination function in a complicated manner precludes the problem from been easily amenable to the minimax treatment which has been useful in linear models. This challenge stimulated the development of the "minave" approach to robust designs and also biased-constrained D-optimality criteria. We first develop these criteria in the context of linear models. The minave design is the design minimizing the average loss function over the contamination neighbourhood (1.7).

The second contribution of the thesis is the application of the minave design approach to designing for generalized linear models with misspecified link functions, in particular, designing for binomial models when the assumed logit link is possibly misspecified. The framework employed is that the true but unknown link function as well as the fitted logit link belong to the generalized family of link functions

$$g(\mu, \lambda) = \log \left[\left\{ \left(\frac{1}{1-\mu} \right)^{\lambda} - 1 \right\} / \lambda \right], \quad (\lambda \geq 0),$$

parameterized by λ. The true link function corresponds to an unknown value of the link parameter that might be different from that which corresponds to the logistic model. The design problem in this circumstance is addressed by casting the problem as a linear predictor misspecification problem using a first order Taylor's approximation to the true link function about the parameter corresponding to the fitted canonical link function. This renders the problem somewhat similar to the original linear predictor misspecification problem but with the form of the contamination known up to an unknown link parameter. In applying the minave approach we assume a range of plausible parameter values for the link parameter. The averaging of the loss function, which is again the average mean squared error of predicted response, is over this range of plausible parameter values.

Lastly, the problem of overdispersion is addressed using the minave approach in a manner similar to the treatment of the link misspecification problem. Again, the overdispersion problem is cast as a linear predictor misspecification problem. However, here the contamination is random with known first and second moments. The averaging in the minave approach is again achieved via averaging over a range of plausible values of the overdispersion parameter.

References

Abdelbasit, K.M. and Plackett, R.L. (1983), "Experimental Design for Binary Data," *Journal of the American Statistical Association*, 78, 90-98.

Biedermann, S., Dette, H. and Pepelyshev, A. (2004), "Some Robust Design Strategies for Percentile Estimation in Binary Response Models," Universität Dortmund, Technical Report 2004-19.

Box, G.E.P., and Draper, N.R. (1959), "A basis for the selection of a response surface design," *Journal of the American Statistical Association*, 54, 622-654.

Chaloner, K. and Larntz, K. (1989), "Optimal Bayesian Design Applied to Logistic Regression Experiments," *Journal of Statistical Inference and Planning*, 21, 191-208.

Chaloner, K. and Verdinelli, I. (1995), "Bayesian Experimental Design: A Review," *Statistical Science*, 10, 273-304.

Chernoff, H. (1953), "Locally Optimal Designs for Estimating Parameters," *Annals of Mathematical Statistics*, 24, 586-602.

Dette, H. and Wong, W.K. (1996), "Optimal Bayesian Designs for Models with Partially Specified Heteroscedastic structure," *Annals of Statistics*, 24(5), 2108-2127.

Dette, H., Haines, L. and Imhof, L. (2003), "Bayesian and maximin optimal designs for heteroscedastic regression models," *The Canadian Journal of Statistics*, 33(2), 221-241.

Fang, Z., and Wiens, D.P. (2000), "Integer-valued, Minimax Robust Designs for Estimation and Extrapolation in Heteroscedastic, Approximately Linear Models," *Journal of the American Statistical Association*, 95, 807-818.

Fedorov, V.V. (1972), *Theory of Optimal Experiments*, Academic Press, New York.

Ford, I, Titterington, D.M. and Kitsos, C.P. (1989), "Recent Advances in Nonlinear Experimental Design," *Technometrics*, 31, 49-60.

Fries, A. and Bhattacharyya, G.K. (1986), "Optimal Design for an Inverse Gaussian Regression Model," *Statistics & Probability Letters* 4, 291-294.

Heo, G., Schmuland, B., and Wiens, D.P. (2001), "Restricted Minimax Robust Designs for Misspecified Regression Models," *The Canadian Journal of Statistics*, 29, 117-128.

Hossain, Md.S. (2002), "Implementation of Robust Designs for Straight Line Regression," Unpublished M.Sc Project Report, University of Alberta, Canada.

Huber, P. (1975), "Robustness and Designs," in: *A Survey of Statistical Design and Linear Models*, ed. J.N. Srivastava, North Holland, Amsterdam, pp. 287-303.

Kiefer (1985), *Jack Carl Kiefer Collected Papers III*, Springer, New York.

Kiefer, J. and Wolfowitz, J. (1960), "The Equivalence of Two Extremum Problems," *Canadian Journal of Mathematics*, 12, 363-366.

King, J. and Wong, W.K. (2000), "Minimax D-Optimal Designs for Logistic Model," Biometrics, 56, 1263-1267.

Khan, M.K. and Yazdi, A.A. (1988), "On D-optimal Designs for Binary Data," *Journal of Statistical Planning and Inference*, 18, 83-91.

Li, K.C. (1984), "Robust Regression Designs when the Design Space Consists of Finitely Many Points," *Annals of Statistics*, 12, 269-282.

Li, K.C., and Notz, W. (1982), "Robust Designs for Nearly Linear Regression," *Journal of Statistical Planning and Inference*, 6, 135-151.

Liu, S.X. and Wiens, D.P. (1997), "Robust Designs for Approximately Polynomial Regression," *Journal of Statistical Planning and Inference*, 64, 369-381.

Marcus, M.B., and Sacks, J. (1976), "Robust Designs for Regression Problems," in: *Statistical Theory and Related Topics II*, ed. S.S. Gupta and D.S. Moore, New York: Academic Press, pp. 245-268.

McCullagh, P. and Nelder, J.A. (1983), *Generalized Linear Models*, Chapman & Hall/CRC.

Minkin, S. (1987), "Optimal Design for Binary Data," *Journal of the American Statistical Association*, 82, 1098-1103.

Minkin, S. (1993), "Experimental Design for Clonogenic Assays in Chemotherapy," *Journal of the American Statistical Association*, 88, 410-420.

Pregibon, D. (1980), "Goodness of Link Selection in Generalized Linear Models," *Applied Statistics*, 29, 15 - 24.

Pesotchinsky, L. (1982), "Optimal Robust Designs: Linear Regression in R^k," *Annals of Statistics*, 10, 511-525.

Pukelsheim, F. (1993), *Optimal Design of Experiments*, New York: Wiley.

Sacks, J. and Ylvisaker, D. (1978), "Linear Estimation for Approximately Linear Models," *Annals of Statistics*, 6, 1122-1137.

Sinha, S. and Wiens, D.P. (2002), "Robust Sequential Designs for Nonlinear Regression," *The Canadian Journal of Statistics*, 30, 601-618.

Sitter, R.R. (1992), "Robust Designs for Binary Data," *Biometrics*, 48, 1145-1155.

Silvey, S.D. (1980), *Optimal Design*, London: Chapman and Hall.

Wiens, D.P. (1990), "Robust, Minimax Designs for Multiple Linear Regression," *Linear Algebra and Its Applications, Second Special Issue on Linear Algebra and Statistics*, 127, 327-340.

Wiens, D.P. (1992), "Minimax Designs for Approximately Linear Regression," *Journal of Statistical Planning and Inference*, 31, 353-371.

Wiens, D.P. (1998), "Minimax Robust Designs and Weights for Approximately Specified Regression Models with Heteroscedastic Errors," *Journal of the American Statistical Association*, 93, 1440-1450.

Wiens, D.P. and Xu, X. (2006), "Robust Prediction and Extrapolation Designs for Misspecified Generalized Linear Regression Models," *Journal of Statistical Inference and Planning*, in press.

Williams, D.A. (1982), "Extra-binomial Variation in Logistic Linear Models," *Applied Statistics*, 31(2), 144-148.

Xu, X. (2006), "Robust Prediction and Extrapolation Designs, with Applications to Accelerated Life Testing," Unpublished Ph.D Dissertation, University of Alberta, Canada.

Yue, R.X. and Hickernell, F.J. (1999), "Robust Designs for Fitting Linear Models with Misspecification," *Statistica Sinica*, 9, 1053-1069.

CHAPTER II

NEW CRITERIA FOR ROBUST INTEGER-VALUED DESIGNS IN LINEAR MODELS

Abstract We investigate the problem of designing for linear regression models, when the assumed model form is only an approximation to an unknown true model, using two novel approaches. The first is based on a notion of averaging of the mean squared error of predictions over a neighbourhood of contaminating functions. The other is based on the usual D-optimal criterion but subject to bias-related constraints in order to ensure robustness to model misspecification. Both approaches are integer-valued constructions in the spirit of Fang and Wiens (2000). Our results are similar to those that have been reported using a minimax approach even though the rationale for the designs presented here are based on the notion of averaging, rather than maximizing, the loss over the contamination space. We also demonstrate the superiority of an integer-valued construction over the continuous designs using specific examples. The designs which protect against model misspecification are clusters of observations about the points that would have been the design points for classical

---

[1] Co-authored with Professor Douglas P. Wiens. To appear in *Computational Statistics and Data Analysis*.

variance-minimizing designs.

2.1 Statistical model

We consider the general regression model with additive errors: $Y = E(Y|\mathbf{x}) + \varepsilon$. Suppose an experimenter is faced with a set $\mathcal{S} = \{\mathbf{x}_i\}_{i=1}^N$ of possible design points from which he is interested in choosing n, not necessarily distinct, points at which to observe Y. The experimenter makes $n_i \geq 0$ observations at $\mathbf{x}_i$ such that $\sum_{i=1}^N n_i = n$. The design problem is to choose $n_1, ..., n_N$ in an optimal manner. Equivalently, the objective is to find an optimal probability distribution $\{p_i\}_{i=1}^N$, with $p_i = n_i/n$, on the design space $\mathcal{S} = \{\mathbf{x}_i\}_{i=1}^N$. The resulting design is said to be *integer valued*.

The experimenter believes that the mean response $E(Y|\mathbf{x})$ may be approximated, but not perfectly described, by a linear combination $\mathbf{z}^T(\mathbf{x})\boldsymbol{\beta}$ of p regressors $\mathbf{z}(\mathbf{x}) = (z_1(\mathbf{x}), ..., z_p(\mathbf{x}))^T$. Since $E(Y|\mathbf{x}) = \mathbf{z}^T(\mathbf{x})\boldsymbol{\beta}$ is just an approximation to the true model, the"best" $\boldsymbol{\beta}_0$ for predicting the mean response is defined to be the minimizer of the average squared error of the approximation:

$$\boldsymbol{\beta}_0 = \arg\min_{\mathbf{t}} \frac{1}{N}\sum_{i=1}^{N}\left(E[Y|\mathbf{x}_i] - \mathbf{z}^T(\mathbf{x}_i)\,\mathbf{t}\right)^2. \tag{2.1}$$

Define $f(\mathbf{x}) = E[Y|\mathbf{x}] - \mathbf{z}^T(\mathbf{x})\boldsymbol{\beta}_0$, so that the model becomes

$$Y_{ij} = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta}_0 + f(\mathbf{x}_i) + \varepsilon_{ij},\ i = 1, ..., N,\ j = 1, ..., n_i, \tag{2.2}$$

where ε_{ij} is the random error associated with the jth observation chosen at the ith design point and $var(\varepsilon_{ij}) = \sigma^2$.

From (2.1), we have

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}(\mathbf{x}_i)f(\mathbf{x}_i) = \mathbf{0}, \tag{2.3}$$

and in order to ensure that the bias in the least squares estimate $\hat{\boldsymbol{\beta}}$ remains within bounds, we place a bound on the misspecification, that is

$$\frac{1}{N}\sum_{i=1}^{N} f^2\left(\mathbf{x}_i\right) \leq \tau^2. \tag{2.4}$$

Let the class of contamination functions $f\left(\mathbf{x}\right)$ satisfying (2.3) and (2.4) be denoted $\mathcal{F}$. We define the loss I as the average mean squared error (AMSE) of $\hat{Y}\left(\mathbf{x}\right) = \mathbf{z}^T\left(\mathbf{x}\right)\hat{\boldsymbol{\beta}}$ as an estimate of $E[Y|\mathbf{x}]$:

$$\begin{aligned} I &= \frac{1}{N}\sum_{i=1}^{N} E\left\{\hat{Y}\left(\mathbf{x}_i\right) - E\left[Y|\mathbf{x}_i\right]\right\}^2 \\ &= \frac{1}{N}\sum_{i=1}^{N}\left(E\left[\hat{Y}\left(\mathbf{x}_i\right)\right] - \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta}\right)^2 \\ &+ \frac{1}{N}\sum_{i=1}^{N} var\left[\hat{Y}\left(\mathbf{x}_i\right)\right] + \frac{1}{N}\sum_{i=1}^{N} f^2\left(\mathbf{x}_i\right) \end{aligned} \tag{2.5}$$

In §2.2 of this article we develop our design criteria. An algorithm to obtain designs minimizing the corresponding loss functions is described in §2.3, with examples and applications in Sections 2.4 and 2.5.

2.2 Loss functions

Fang and Wiens (2000) used a minimax approach to construct integer-valued designs. The optimal design in the minimax sense is the design that minimizes the maximum, over the misspecification neighbourhood $\mathcal{F}$, value of the loss. The minimax approach aims to obtain the best design for the worst possible case of model misspecification. Here we introduce new criteria for robust designs which may have more intuitively appeal to practitioners. Rather than minimizing the maximum loss we instead choose the design which minimizes the *average* value of the loss over the misspecification neighbourhood. The averaging requires a parameterization of $\mathcal{F}$ which is carried out in §2.2.1 below. This approach can be seen as a generalization of Läuter's (1974, 1976) approach. Läuter accommodated model uncertainty in the choice of design by averaging design criterion functions over a finite set of plausible models. Here we have an infinite set of plausible models defined by (2.2), (2.3) and (2.4). While Läuter's criterion is based on variance only, in the spirit of Box and Draper (1959) we base our design criteria on possible bias engendered by the model misspecification as well as on variance.

2.2.1 "Minave" Mean Squared Error model-robust design criteria

Given the misspecification neighbourhood, we seek integer-valued designs that minimize the average (over $\mathcal{F}$) value of the loss. Let $\{p_i = n_i/n\}_{i=1}^{N}$ be an integer-valued design on $\mathcal{S}$, $\mathbf{P}$ the $N \times N$ diagonal matrix with diagonal elements $\{p_i\}$, $\mathbf{Z}$ the $N \times p$ matrix, assumed to be of full rank, with rows $\mathbf{z}^T(\mathbf{x}_1), ..., \mathbf{z}^T(\mathbf{x}_N)$. Define $\mathbf{f} = (f(\mathbf{x}_1), ..., f(\mathbf{x}_N))^T$. In this notation, the average mean squared error defined in (2.5) can be written as

$$I = \frac{1}{N}\left\{\frac{\sigma^2}{n} tr\left[\left(\mathbf{Z}^T\mathbf{P}\mathbf{Z}\right)^{-1}\mathbf{Z}^T\mathbf{Z}\right] + \mathbf{f}^T\mathbf{P}\mathbf{Z}\left(\mathbf{Z}^T\mathbf{P}\mathbf{Z}\right)^{-1}\mathbf{Z}^T\mathbf{Z}\left(\mathbf{Z}^T\mathbf{P}\mathbf{Z}\right)^{-1}\mathbf{Z}^T\mathbf{P}\mathbf{f} + \mathbf{f}^T\mathbf{f}\right\}.$$

Note that we are assuming that the design is feasible for the full parameter vector $\boldsymbol{\beta}$, or equivalently that it has a minimum of p distinct support points $\mathbf{x}_i$ in $\mathcal{S}$ such that the vectors $\mathbf{z}(\mathbf{x}_i)$ are linearly independent. This implies the non-singularity of $\mathbf{Z}^T\mathbf{P}\mathbf{Z}$.

We obtain the singular value decomposition of $\mathbf{Z}$, $\mathbf{Z} = \mathbf{U}_{N\times p}\boldsymbol{\Lambda}_{p\times p}\mathbf{V}_{p\times p}^T$ say, with $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}_p$ and $\boldsymbol{\Lambda}$ diagonal and invertible. The $N \times p$ matrix $\mathbf{U}$ is augmented by $\tilde{\mathbf{U}}_{N\times(N-p)}$ such that $\left[\mathbf{U}\vdots\tilde{\mathbf{U}}\right]_{N\times N}$ is orthogonal. Then by (2.3) and (2.4), we have that there is an $N-p\times 1$ vector $\mathbf{c}$, with $\|\mathbf{c}\|\leq 1$, satisfying $\mathbf{f} = \tau\sqrt{N}\tilde{\mathbf{U}}\mathbf{c}$, and then

$$I = \frac{1}{N}\left\{\frac{\sigma^2}{n}tr\left[\left(\mathbf{U}^T\mathbf{P}\mathbf{U}\right)^{-1}\right] + \tau^2 N tr\left[\tilde{\mathbf{U}}^T\mathbf{P}\mathbf{U}\left(\mathbf{U}^T\mathbf{P}\mathbf{U}\right)^{-2}\mathbf{U}^T\mathbf{P}\tilde{\mathbf{U}}\mathbf{c}\mathbf{c}^T\right] + \tau^2 N\mathbf{c}^T\mathbf{c}\right\}.$$

For details of this development see Fang and Wiens (2000). We define our design criterion as I, with $\mathbf{f}$ integrated over $\mathbf{c}$:

$$I_{ave} = \frac{\sigma^2}{nN}tr\left[\left(\mathbf{U}^T\mathbf{P}\mathbf{U}\right)^{-1}\right] + \tau^2\int_{\|\mathbf{c}\|\leq 1}\left(tr\left[\tilde{\mathbf{U}}^T\mathbf{P}\mathbf{U}\left(\mathbf{U}^T\mathbf{P}\mathbf{U}\right)^{-2}\mathbf{U}^T\mathbf{P}\tilde{\mathbf{U}}\mathbf{c}\mathbf{c}^T\right] + \mathbf{c}^T\mathbf{c}\right)d\mathbf{c}.$$

Theorem 2.1 *Define*

$$\kappa_{N,p} = \frac{\pi^{(N-p)/2}}{\left(\frac{N-p}{2}+1\right)\Gamma\left(\frac{N-p}{2}\right)} = \int_{\|\mathbf{c}\|\leq 1}\mathbf{c}^T\mathbf{c}d\mathbf{c}.$$

The average of I, the Average Mean Squared Error over the misspecification neighbourhood $\mathcal{F}$, is given by $I_{ave} = \left(\frac{\sigma^2}{n} + \tau^2\kappa_{N,p}\right)\mathcal{L}_{ave}$, where

$$\mathcal{L}_{ave} = \rho\frac{tr\left[\left(\mathbf{U}^T\mathbf{P}\mathbf{U}\right)^{-1}\right]}{N} + (1-\rho)\left(1 + \frac{tr\left[\left(\mathbf{U}^T\mathbf{P}\mathbf{U}\right)^{-2}\left(\mathbf{U}^T\mathbf{P}^2\mathbf{U}\right)\right] - p}{N-p}\right) \qquad (2.6)$$

for $\rho = \frac{\sigma^2}{n}/\left(\frac{\sigma^2}{n} + \tau^2\kappa_{N,p}\right)$.

Proof: Note that $\int_{\|\mathbf{c}\|\leq 1}\mathbf{c}\mathbf{c}^T d\mathbf{c} = \frac{\kappa_{N,p}}{N-p}\mathbf{I}_{N-p}$. We then calculate that

$$\begin{aligned} I_{ave} &= \frac{\sigma^2}{nN}tr\left[\left(\mathbf{U}^T\mathbf{P}\mathbf{U}\right)^{-1}\right] + \tau^2 tr\left[\tilde{\mathbf{U}}^T\mathbf{P}\mathbf{U}\left(\mathbf{U}^T\mathbf{P}\mathbf{U}\right)^{-2}\mathbf{U}^T\mathbf{P}\tilde{\mathbf{U}}\int_{\|\mathbf{c}\|\leq 1}\mathbf{c}\mathbf{c}^T d\mathbf{c}\right] + \tau^2\kappa_{N,p} \\ &= \frac{\sigma^2}{nN}tr\left[\left(\mathbf{U}^T\mathbf{P}\mathbf{U}\right)^{-1}\right] + \tau^2\frac{\kappa_{N,p}}{N-p}tr\left[\mathbf{P}\mathbf{U}\left(\mathbf{U}^T\mathbf{P}\mathbf{U}\right)^{-2}\mathbf{U}^T\mathbf{P}\left(\mathbf{I}_N - \mathbf{U}\mathbf{U}^T\right)\right] + \tau^2\kappa_{N,p}, \end{aligned}$$

which reduces to (2.6).

Since $\left(\frac{\sigma^2}{n} + \tau^2\kappa_{N,p}\right)$ does not depend on the design, we use the loss $\mathcal{L}_{ave}$ for the design construction. We call $\frac{1}{N}tr\left[\left(\mathbf{U}^T\mathbf{P}\mathbf{U}\right)^{-1}\right]$ and $1 + \frac{tr\left[\left(\mathbf{U}^T\mathbf{P}\mathbf{U}\right)^{-2}\left(\mathbf{U}^T\mathbf{P}^2\mathbf{U}\right)\right]-p}{N-p}$ the variance and bias discrepancies respectively; $\mathcal{L}_{ave}$ is a weighted average of them. The term ρ may be interpreted as the relative contribution of variance to mean squared error. Values of ρ near 1 mean that variance is considered much more important than bias in the mind of the experimenter, while values of $1-\rho$ near 1 indicate a correspondingly large concern with bias. Thus ρ can be understood as the prior belief of the experimenter as to the nature of the true response function. Our "minave" design is the design that minimizes $\mathcal{L}_{ave}$ for a given value of ρ.

2.2.2 Bias-Constrained D-optimal model-robust design criteria

The design criteria we consider here are motivated by a problem we encountered in the context of robust designs for generalized linear models. In the presence of model misspecification, the asymptotic covariance matrix of estimated model parameters in generalized linear models does not in general simply equal the inverse of the information matrix. The asymptotic covariance matrix is rather a function of the information matrix and the variance of the response Y under the true model. Thus robust designs for generalized linear models would seem to require not only knowledge of the model parameters as in classical designs for generalized linear models but also knowledge of the variance of the response under the true but unknown model. The approach reported here is motivated by a requirement to generate robust designs without assuming this latter knowledge.

Here we develop and apply this approach in the context of linear models; considerations for generalized linear models are reported in Adewale and Wiens (2005). We explore designs based on constrained maximization of the determinant of the information matrix of the fitted model parameters. Robustness is achieved via the constraint which represents some bound on a measure of the bias engendered by the model misspecification. Montepiedra and Fedorov (1995) proposed similar criteria using convex design theory when the true model is known specifically but the fitted model is, for the sake of parsimony, a simplified version of the true model. Their consideration is in the same setting as that of Box and Draper (1959) but instead of using a function of the mean squared error as loss function they worked with constrained variance or bias, this in order to keep the mathematics within the context of convex theory where the criteria would be amenable to Kiefer's Equivalence Theorem. Our construction differs in three senses: the true model is assumed unknown, we focus on finite design space and our designs are integer-valued. We investigate two kinds of constraints:

1. bounding the norm of the bias in the estimated regression parameter vector; and
2. bounding the average (over $\mathcal{F}$) of the average norm (over $\mathcal{S}$) of the bias in the predicted response.

This first constraint might be suitable when the focus of the design is on parameter estimation while the second would be recommended when the interest is in prediction.

The bias in $\hat{\boldsymbol{\beta}}$ is calculated as

$$bias\left(\hat{\boldsymbol{\beta}}\right) = E\left[\hat{\boldsymbol{\beta}}\right] - \boldsymbol{\beta}_0 = \left(\mathbf{Z}^T\mathbf{P}\mathbf{Z}\right)^{-1}\mathbf{Z}^T\mathbf{P}\mathbf{f}.$$

A development similar to that preceding Theorem 2.1 yields that the average squared norm of the bias in $\hat{\boldsymbol{\beta}}$ is

$$aver_{\mathcal{F}}\left(\left\|bias\left(\hat{\boldsymbol{\beta}}\right)\right\|^{2}\right) = aver_{\mathcal{F}}\left(\mathbf{f}^{T}\mathbf{PZ}\left(\mathbf{Z}^{T}\mathbf{PZ}\right)^{-2}\mathbf{Z}^{T}\mathbf{Pf}\right)$$

and the average squared norm of the bias in the predicted response is

$$aver_{\mathcal{F}}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|bias\left(\mathbf{z}^{T}\left(\mathbf{x}_{i}\right)\hat{\boldsymbol{\beta}}\right)\right\|^{2}\right) = \frac{1}{N}aver_{\mathcal{F}}\left\{\mathbf{f}^{T}\mathbf{PZ}\left(\mathbf{Z}^{T}\mathbf{PZ}\right)^{-1}\mathbf{Z}^{T}\mathbf{Z}\left(\mathbf{Z}^{T}\mathbf{PZ}\right)^{-1}\mathbf{Z}^{T}\mathbf{Pf} + \mathbf{f}^{\mathbf{T}}\mathbf{f}\right\}$$

In each case we average over $\mathcal{F}$ through the vector $\mathbf{c}$ and obtain the following results.

Theorem 2.2 *The average squared norm of the bias in $\hat{\boldsymbol{\beta}}$ over the misspecification neighbourhood $\mathcal{F}$ is given by*

$$aver_{\mathcal{F}}\left(\left\|bias\left(\hat{\boldsymbol{\beta}}\right)\right\|^{2}\right) = \frac{\kappa_{N,p}\tau^{2}N}{N-p}tr\left\{\left[\left(\mathbf{U}^{T}\mathbf{PU}\right)^{-1}\left(\mathbf{U}^{T}\mathbf{P}^{2}\mathbf{U}\right)\left(\mathbf{U}^{T}\mathbf{PU}\right)^{-1}-\mathbf{I}\right]\boldsymbol{\Lambda}^{-2}\right\}.$$

Theorem 2.3 *The average squared norm of bias in predicted response over $\mathcal{F}$ and $\mathcal{S}$ is given by*

$$aver_{\mathcal{F}}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|bias\left(\mathbf{z}^{T}\left(\mathbf{x}_{i}\right)\hat{\boldsymbol{\beta}}\right)\right\|^{2}\right) = \kappa_{N,p}\tau^{2}\left(1+\frac{tr\left[\left(\mathbf{U}^{T}\mathbf{PU}\right)^{-2}\left(\mathbf{U}^{T}\mathbf{P}^{2}\mathbf{U}\right)\right]-p}{N-p}\right).$$

The classical D-optimal design is the design that maximizes the determinant of the information matrix of the model parameters or, equivalently, the design that minimizes the determinant of the covariance matrix of the model parameters. Here, our robust design is the design that maximizes the determinant of the information matrix (divided by $\det\left(\boldsymbol{\Lambda}^{2}\right)$, which does not depend on the design) subject to either of the two bias-related constraints. That is:

Criterion 1: Robust Estimation. Choose $\mathbf{P}$ so as to maximize det $\left\{\mathbf{U}^T\mathbf{P}\mathbf{U}\right\}$ subject to

$$tr\left\{\left[\left(\mathbf{U}^T\mathbf{P}\mathbf{U}\right)^{-1}\left(\mathbf{U}^T\mathbf{P}^2\mathbf{U}\right)\left(\mathbf{U}^T\mathbf{P}\mathbf{U}\right)^{-1}-\mathbf{I}\right]\mathbf{\Lambda}^{-2}\right\}\leq\alpha. \qquad (2.7)$$

Criterion 2: Robust Prediction. Choose $\mathbf{P}$ so as to maximize det $\left\{\mathbf{U}^T\mathbf{P}\mathbf{U}\right\}$ subject to

$$tr\left\{\left(\mathbf{U}^T\mathbf{P}\mathbf{U}\right)^{-2}\left(\mathbf{U}^T\mathbf{P}^2\mathbf{U}\right)\right\}\leq\beta. \qquad (2.8)$$

The resulting design from the estimation criterion depends on α, while that of the prediction criterion depends on β. For values of α and β exceeding those attained by the classical D-optimal design, the 'robust' and classical designs coincide.

2.3 Numerical algorithms

2.3.1 Simulated annealing algorithm for Minave designs

The optimization problem for Minave designs is an unconstrained nonlinear integer optimization problem. We consider models with p regressors $(z_1(x), ..., z_p(x))^T$ where $x\in[-1,1]$. Given the desired number of observations (n) to be taken and the number of points in the design space (N), we seek designs that minimize the Minave loss function (2.6). Our algorithm accommodates all (n, N) combinations. If the problem of interest demands a design which is symmetric about the origin, (n, N) must be chosen such that symmetry is possible. For instance, symmetric designs would not be obtainable when N is even and n is odd. We take $\mathcal{S}$ to be the set $\left\{x_i=-1+\frac{2(i-1)}{N-1}\right\}_{i=1}^{N}$ of equally spaced points in $[-1,1]$.

Simulated annealing is employed to search for optimal designs. The simulated annealing algorithm seeks to assign integers $n_i\geq 0$ to each of the design points

x_i in such a way that $\mathcal{L}_{ave}$ is a minimum. Simulated annealing is a direct search optimization algorithm which has been quite successful at finding the global extremum of a function, possibly non-smooth, that has many local extrema. The algorithm is a biased random walk consisting of three steps. The first step is a specification of the initial state, $\mathbf{n}_0$, of the process based on which the corresponding initial value of the objective function $\mathcal{L}_{ave}(\mathbf{n}_0)$ is calculated. The second step is the random choice of the next state of the process from the optimization space. The last step is a prescription of the basis of acceptance or rejection of the new state.

For the algorithm to be successful, each step has to be empirically tailored to the context of the problem at hand. We randomly choose p points in $\mathcal{S}$ and the n observations are randomly allocated to these p locations. This constitutes the initial state of the simulated annealing process. When the interest is in symmetric designs we randomly pick $\lceil p/2 \rceil$ points of the points of $\mathcal{S}$ in $[-1, 0)$ and randomly allocate $\lceil n/2 \rceil$ observations to these points. If n is odd then N has to be odd for symmetry, in this case replace $\lceil n/2 \rceil$ by $\lfloor n/2 \rfloor$ and assign one observation to the point $\{0\}$. A symmetric initial design is obtained by assigning the number of observations for locations in $[-1, 0)$ to their corresponding mirror image about the origin. Fang and Wiens (2000) assumed that one of (n, N) is a multiple of the other and then chose the initial design to be as uniform as possible. In the examples presented, we considered this approach to choosing the initial design as well. With either choice of initial design the algorithm converges to the same design.

To generate a new design, we perturb the current state as described in Fang and

Wiens (2000). Define $\mathbf{v}$ to be the $N \times 1$ current allocation vector. For symmetric designs redefine $\mathbf{v}$ to be the $[N/2] \times 1$ vector consisting of the initial segment $\left(n_1, ..., n_{[N/2]}\right)$ of the current allocation vector. Let $J_+ = \{i|v_i > 0\}$, $J_0 = \{i|v_i = 0\}$ with cardinalities $j_+ \geq 1$ and j_0. If $j_+ \geq 2$, generate a Bernoulli random variable

$$B = \begin{cases} 1 & \text{with probability } j_0/\left(j_0 + j_+\right), \\ 0 & \text{with probability } j_+/\left(j_0 + j_+\right), \end{cases}$$

choose two indices (t_1, t_2) from J_+, at random without replacement, choose an index t_0 from J_0, at random and modify the selected components of $\mathbf{v}$ as follows:

$$v_{t_0} = v_{t_0} + B,\ v_{t_1} = v_{t_1} - 1 \text{ and } v_{t_2} = v_{t_2} + 1 - B. \tag{2.9}$$

If $j_+ = 1$, choose t_0 from J_0 at random, let t_1 be the index in the singleton set J_+, and then replace (2.9) by

$$v_{t_0} = v_{t_0} + 1,\ v_{t_1} = v_{t_1} - 1.$$

This completes the perturbation scheme for general designs. For symmetric designs, we complete the scheme as follows. If N is even, then let $\mathbf{n} = (n_1, ..., n_N) = \left(v_1, ..., v_{N/2}, v_{N/2}, ..., v_1\right)$. If N is odd, then generate a uniform random variable u. If $u < 1/N$, with probability $1/2$ increase $n_{[N/2]+1}$ by 2 then randomly and symmetrically reduce the remaining n_i by 2; with probability $1/2$ reduce $n_{[N/2]+1}$ by 2 then randomly and symmetrically increase the remaining n_i. This step is omitted if $n_{[N/2]+1} < 2$. We then construct $\mathbf{n}$ as described above, with the inclusion of the new frequency $n_{[N/2]+1}$.

The value of the loss $\mathcal{L}_{ave} = \mathcal{L}_{ave}\left(\tilde{\mathbf{n}}\right)$ is evaluated at the new state $\tilde{\mathbf{n}}$ and the state is accepted with probability π, defined as

$$\pi = \begin{cases} 1 & \text{if } \Delta\mathcal{L}_{ave} \leq 0, \\ \exp\left(-\Delta\mathcal{L}_{ave}/T\right) & \text{if } \Delta\mathcal{L}_{ave} \geq 0, \end{cases}$$

where $\Delta\mathcal{L}_{ave} = \mathcal{L}_{ave}(\tilde{\mathbf{n}}) - \mathcal{L}_{ave}(\mathbf{n})$. Thus a favourable state ($\Delta\mathcal{L}_{ave} \leq 0$) is accepted with certainty and an unfavourable state is accepted according to a separate Bernoulli experiment with success probability $\exp\left(-\Delta\mathcal{L}_{ave}/T\right)$. We choose T such that initially the inequality $.5 < \exp\left(-\Delta\mathcal{L}_{ave}/T\right) < .9$ is satisfied; this follows a suggestion from Bohachevsky, Johnson and Stein (1986). As long as $\exp\left(-\Delta\mathcal{L}_{ave}/T\right) > 0$ an unfavourable state could be accepted, thus providing the possibility of the path leading out of local minima. To ensure that the process settles at a global minimum we progressively decrease T. Fang and Wiens (2000) decrease T by a factor of .9 after each 100 iterations. We found that it is better to adapt the change of T to the specific problem of interest. In some cases we were able to decrease T by a factor of .95 after every 20^{th} iteration.

In Figure 1 we present the simulated annealing trajectory, for one of the examples of the next section, to illustrate our algorithm. As seen in this plot the simulated annealing process walks from one design to the next irrespective of whether the step taken is progressive (when it moves to a design with a lesser loss) or detrimental (when it moves to a design with a higher loss). The ability to accept detrimental steps prevents the process from being trapped in local extrema. After a large number of iterations, in this case 1500, it is expected that the algorithm has settled at a design with (near) minimum loss. In the examples below we satisfied ourselves of the convergence to a minimum by varying the number of iterations, the factor by which

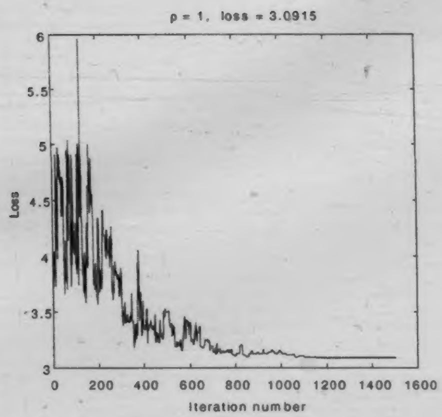

Figure 1: Simulated annealing trajectory for cubic regression design with $N = 40$, $n = 20$ and $\rho = 1$.

the temperature parameter is decreased and by running the algorithm many times.

2.3.2 Simulated annealing algorithm for Bias-Constrained D-optimal designs

Designing using our bias-constrained criteria is an integer optimization problem with a nonlinear constraint. The approach taken is to convert the problem into an unconstrained integer optimization problem such that the preceding algorithm for unconstrained designs becomes applicable. We achieved this by adding a penalty function to the objective function. Since the optimal design of interest is that which maximizes the objective function, we choose a negative-valued penalty function. Thus the constrained optimization problems (2.7) and (2.8) are converted to

$$\text{Maximize } \det\left\{\mathbf{U}^T\mathbf{P}\mathbf{U}\right\} + g\left(x\right) I_{[x>\alpha]}, \tag{2.10}$$

$$\text{where } x = tr\left\{\left[\left(\mathbf{U}^T\mathbf{P}\mathbf{U}\right)^{-1}\left(\mathbf{U}^T\mathbf{P}^2\mathbf{U}\right)\left(\mathbf{U}^T\mathbf{P}\mathbf{U}\right)^{-1} - \mathbf{I}\right]\mathbf{\Lambda}^{-2}\right\}$$

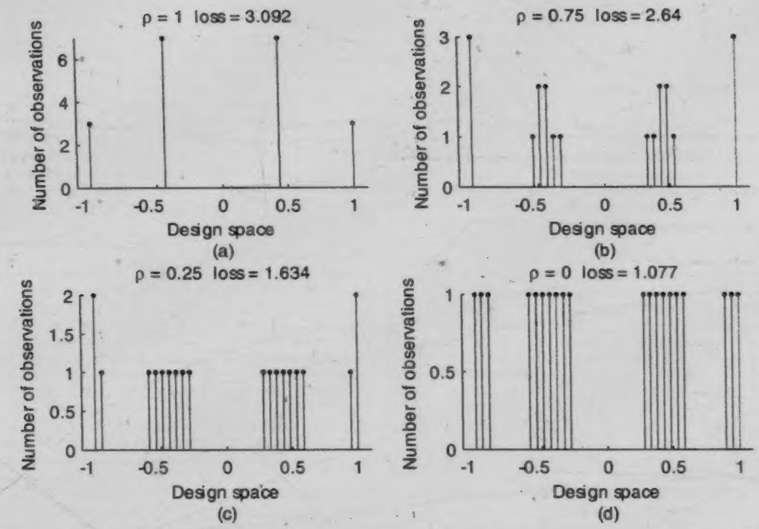

Figure 2: Integer-valued designs for cubic regression with $N = 40$ and $n = 20$ for (a) $\rho = 1$, (b) $\rho = .75$, (c) $\rho = .25$ and (d) $\rho = 0$.

and

$$\text{Maximize } \det\left\{\mathbf{U}^T\mathbf{P}\mathbf{U}\right\} + g(x)\, I_{[x>\beta]}, \qquad (2.11)$$
$$\text{where } x = tr\left\{\left(\mathbf{U}^T\mathbf{P}\mathbf{U}\right)^{-2}\left(\mathbf{U}^T\mathbf{P}^2\mathbf{U}\right)\right\}$$

respectively, where $g(\cdot)$ is a negative-valued penalty function and $I_{[\cdot]}$ is an indicator function. Given $g(\cdot)$, the optimization problems become amenable to the algorithm we used in the preceding section for unconstrained integer optimization. We found that the algorithm was generally unsuccessful when the penalty function is chosen to be a constant, that is when $g(x) = -c$, where c is a positive real number. The choice of $g(\cdot)$ that we have found useful for solving (2.10) and (2.11) is instead $g(x) = -c\cdot x$, $c > 0$. In the examples we reported we chose $c = 1$; other values of c worked equally well.

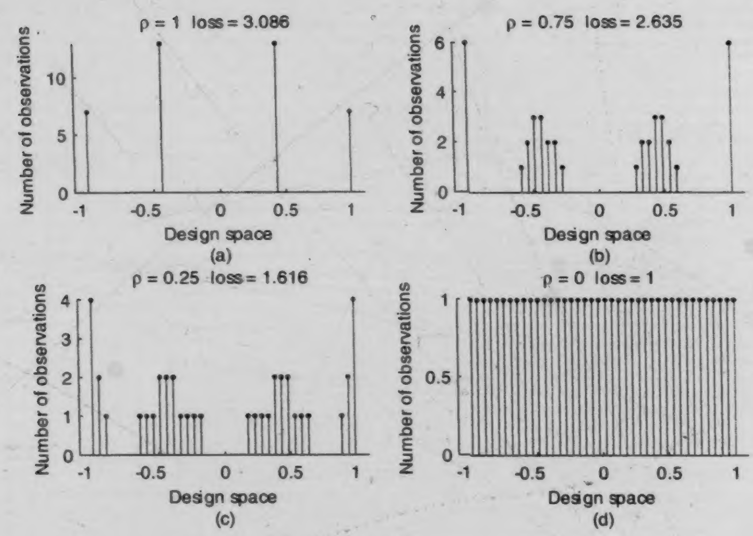

Figure 3: Integer-valued designs for cubic regression with $N = 40$ and $n = 40$ for (a) $\rho = 1$, (b) $\rho = .75$, (c) $\rho = .25$ and (d) $\rho = 0$.

2.4 Examples: Polynomial regression

Example 1. For the purpose of comparison with the minimax designs of Fang and Wiens (2000), we consider approximate cubic regression. For $\rho = 1$ (so that $\mathcal{L}_{ave}$ coincides with the average variance of the predicted values), $n = 20$ and $N = 40$ we obtain the design placing 3 of the 20 observations at each of ±1 and 7 at each of ±.436. This agrees with the result of Fang and Wiens (2000) when the value of their design parameter v, measuring the importance of variance relative to bias, is taken to be very large. See Figure 2 for this and other obtained designs, all for $N = 40$. In contrast, when $\rho = 0$ the designs are as uniform as is allowed by the (n, N) combination.

In Figure 3 and 4, we present designs for $n = 40$ and $n = 60$, respectively, for values of ρ ranging from 1 (all-variance design) to 0 (all-bias design). A comparison of the

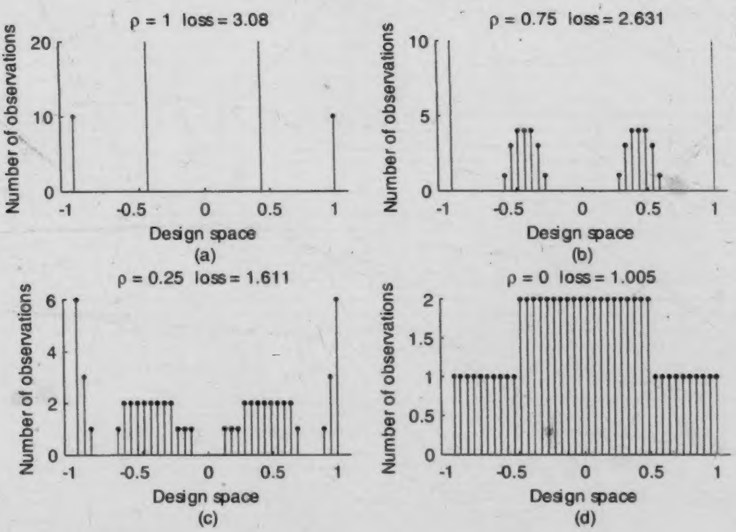

Figure 4: Integer-valued designs for cubic regression with $N = 40$ and $n = 60$ for (a) $\rho = 1$, (b) $\rho = .75$, (c) $\rho = .25$ and (d) $\rho = 0$.

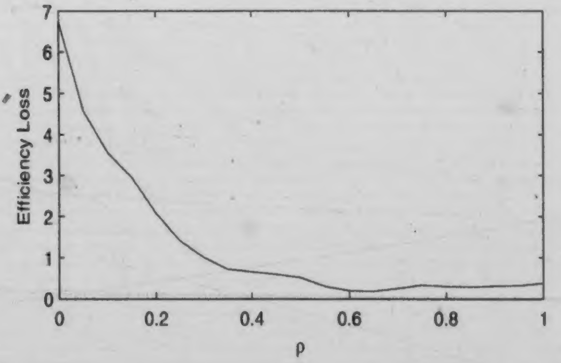

Figure 5: Plot of efficiency loss (versus ρ) due to mere replications of the design points for $n = 20$ as opposed to an explicit construction of the designs for $n = 60$.

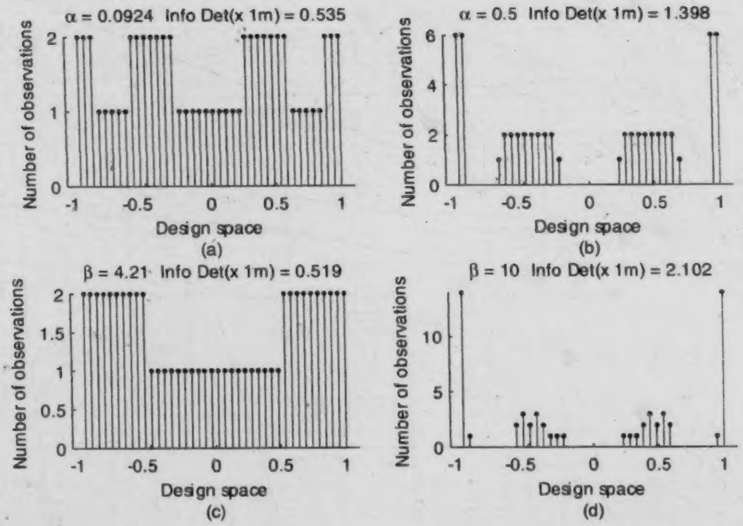

Figure 6: Robust designs for parameter estimation with (a) $\alpha = 0.0924$, (b) $\alpha = 0.5$ and prediction with (c) $\beta = 4.2068$ and (d) $\beta = 10$.

results for cases when $n < N$, $n = N$ and $n > N$ underscores the advantage of an exact integer-valued design over an approximate continuous design. Having obtained the design: $\pm 1\,(3)$, $\pm .436\,(7)$ for $n = 20$ and $\rho = 1$, without an explicit investigation of the case when $n = 40$ and $\rho = 1$, one would have been tempted to suggest the design: $\pm 1\,(6)$, $\pm .436\,(14)$ which would have been 99.8% efficient still because the design points are the same in this case. In the same vein, having obtained the design for $n = 20$ and $\rho = .2$ to obtain the design for $n = 60$ and $\rho = .2$ one would probably multiply the number of observations taken at each of the design points obtained for $n = 20$ by 3. However, an explicit construction of the design for $n = 60$ and $\rho = .2$ resulted in a different design. We found that a naive replication would have resulted in about 2% loss in efficiency. The efficiency loss is much more here because the design for $n = 60$ and $\rho = .2$ includes other points that are not included in the design for

$n = 20$ and $\rho = .2$. Figure 5 is a plot of efficiency loss due to a naive replication of the design points for $n = 20$ as opposed to an explicit construction of design for $n = 60$ for varying values of our design parameter, ρ. The highest loss of efficiency occurs when $\rho = 0$ (so that $\mathcal{L}_{ave}$ coincides with the average bias of the predicted values). This comparison emphasizes the need to be cautious in implementing approximate continuous designs and the superiority of an integer-valued (exact) construction for specific applications.

The resulting designs from the two bias-constrained criteria coincide with the classical D-optimal design, that is, the design that maximizes the determinant of the information matrix, for sufficiently large values of α or β. The classical D-optimal design is the design that takes the assumed model to be exact. Gaffke (1987), Huang (1987), Haines (1987) and Chen and Huang (2000) have constructed exact D-optimal designs, with 'optimality' holding when the fitted model is exactly correct. In contrast, the constraints imposed here protects against misspecifications in the assumed model within a finite design space framework. The least attainable α (resp., β), say α_0 (β_0), corresponds to the value of the constraint for an all-bias design. When n is a multiple of N, α_0 (β_0) corresponds to the bias of the, minimum bias, uniform design. For the estimation criterion it follows from (2.3) that $\alpha_0 = 0$ and for the prediction criterion $\beta_0 = p$, the number of parameters in the model. When n is not a multiple of N, the design with minimum bias has $\alpha_0 > 0$ for the estimation criterion and $\beta_0 > p$ for the prediction criterion.

Example 2: Bias-Constrained D-optimal Designs. We again use the cubic regression model to illustrate these criteria. Guest (1958) and Hoel (1958) independently reported an approach based on zeros of derivatives of Legendre polynomials for obtaining the D-optimal support points for polynomials regression. Using this approach, Pukelsheim (1993, p. 218) presented the D-optimal design over the interval $[-1,1]$ for cubic regression as the design placing equal mass at each of $\pm.447$, ±1. For large α or β, whichever is appropriate, our simulated annealing algorithm results in a design taking an equal number of observations at each of $\pm.436$, ±1. Our algorithm attains the closest approximation to that given above since the points $\pm.436$ are the nearest, in our design space, to $\pm.447$. In panel (a) of Figure 6 we present robust parameter estimation design when $N = 40$, $n = 60$ with α set to a value ($\alpha = .0924$) just slightly greater than the least admissible α ($\alpha_0 = 0.0923$), for this setting. A corresponding design for the robust prediction design criterion is presented in Figure 6(c), for this $\beta_0 = 4.2067$ and $\beta = 4.2068$. Design for another value of α (parameter estimation design) and β (prediction design) are presented in panels (b) and (d) respectively. Notable is the fact that each constraint shapes the designs differently.

2.5 Designs for multiple regression

The algebra of Sections 2.1 and 2.2 applies to general regression models. However, the simulated annealing algorithm described in §2.3 may not be readily useful in situations where there are 2 or more independent variables. This is because the design space is then multidimensional, raising the usual problems related to the curse of dimensionality. Here we consider first and second order multiple regression models.

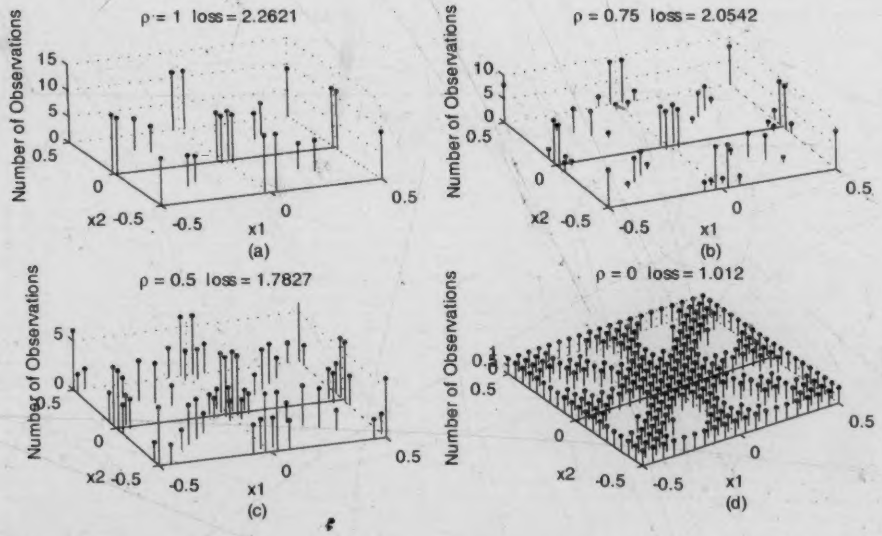

Figure 7: Minave designs for partial second-order model (a) $\rho = 1$, (b) $\rho = .75$, (c) $\rho = .5$, (d) $\rho = 0$.

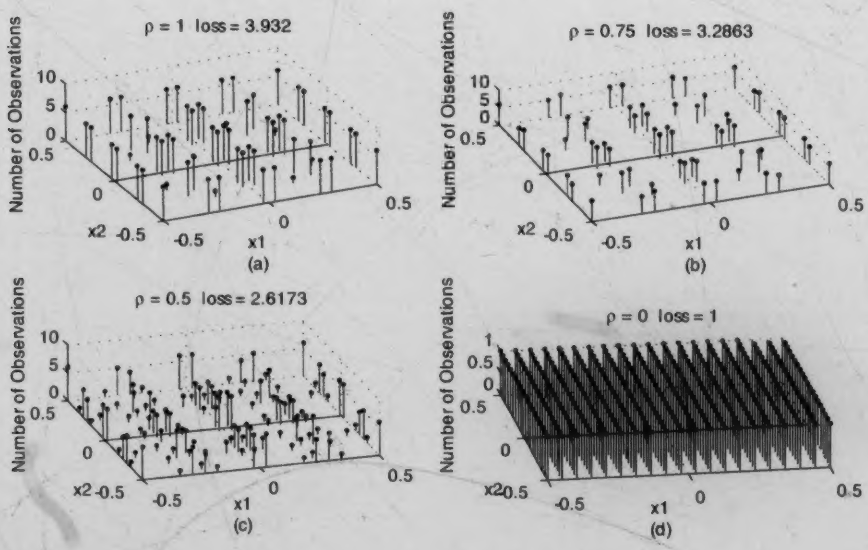

Figure 8: Minave designs for full second-order model (a) $\rho = 1$, (b) $\rho = .75$, (c) $\rho = .5$, (d) $\rho = 0$.

Table 1: Minave Designs for Partial Second-order Model.

ρ	(Design points; frequencies) in half-quadrant $0 \leq x_1 \leq x_2 \leq 1/2$	Loss
1	(.0263, .0263; 9) , (.2368, .2368; 5) , (.2895, .2895; 6) , (.5, .0263; 11) , (.5, .5; 9)	2.2621
.75	(.0263, .0263; 8) , (.1842, .1842; 1) , (.2368, .2368; 5) , (.2895, .2895; 5) , (.3421, .3421; 1) , (.4474, .0263; 1) , (.5, .0263; 9) , (.5, .0789; 2) , (.5, .5; 8)	2.0542
.5	(.0263, .0263; 6) , (.0789, .0263; 2) , (.1842, .1842; 2) , (.2368, .2368; 4) , (.2895, .2895; 4) , (.3421, .3421; 2) , (.4474, .0263; 3) , (.5, .0263; 6) , (.5, .0789; 3) , (.5, .4474; 2) , (0.5, .5; 6)	1.7827
0	(.0263, .0263; 1) , (.0789, .0263; 1) , (.0789, .0789; 1) , (.1316, .0263; 1) , (.1316, .0789; 1) , (0.1316, .1316; 1) , (.1842, .0263; 1) , (.1842, .0789; 1) , (.1842, .1316; 1) , (.1842, .1842; 1) , (.2368, .0263; 1) , (.2368, .2368; 1) , (.2895, .0263; 1) , (.2895, .1842; 1) , (.2895, .2368; 1) , (.2895, .2895; 1) , (.3421, .2895; 1) , (.3421, .3421; 1) , (.3947, .0263; 1) , (.3947, .0789; 1) (.3947, .3421; 1) , (.3947, .3947; 1) , (.4474, .0263; 1) , (.4474, .0789; 1) (.4474, .1316; 1) , (.4474, .3421; 1) , (.4474, .4474; 1) , (.5, .0263; 1) , (.5, .0789; 1) , (.5, .1316; 1) , (.5, .1842; 1) , (.5, .5; 1)	1.0120

Table 2: Minave Designs for Full Second-order Model.

ρ	(Design points; frequencies) in half-quadrant $0 \leq x_1 \leq x_2 \leq 1/2$	Loss
1	(.0263, .0263; 6), (.2368, .0263; 6),(.2368, .2368; 6), (.2895, .0263; 6),(.2895, .2895; 6),(.3421, .1842; 1), (.5, .0263; 6), (.5, .2368; 6),(.5, .2895; 6), (.5, .5; 6)	3.9320
.75	(.0263, .0263; 7), (.2368, .0263; 5), (.2368, .2368; 5), (.2895, .0263; 6),(.2895, .2895; 6), (.3421, .1842; 3), (.5, .0263; 6), (.5, .2368; 6), (.5, .2895; 5),(.5, .5; 6)	3.2863
.5	(.0263, .0263; 6), (.1842, .0263; 1), (.1842, .1842; 1), (.2368, .0263; 5),(.2368, .2368; 5), (.2895, .0263; 4), (.2895, .0789; 1), (.2895, .2368; 1),(.2895, .2895; 4), (.3421, .0263; 1),(.3421, .1842; 1), (.3421, 3421; 1), (.4474, .2368; 2),(.5, .0263; 6), (.5, .1842; 1), (.5, .2368; 4), (.5, .2895; 4),(.5, .3421; 1), (.5, .5; 6)	2.6173
0	{*all points on the design space with one observation taken at point*}	1.0000

Suppose $\mathbf{x} = (x_1, ..., x_q)^T$ and $\mathbf{z}(\mathbf{x})$ has elements $1, x_1, ..., x_q$ and possibly second order terms $x_i x_j$ $(1 \leq i \leq j \leq q)$. We define the design space as the $q-$fold Cartesian product $\mathcal{S} = \mathcal{S}_1 \times \cdots \times \mathcal{S}_q$, where each $\mathcal{S}_j$ has N_0 equally spaced points in $\left[-\frac{1}{2}, \frac{1}{2}\right]$. Thus, the design space consists of $N = N_0^q$ points. If, for instance, $N_0 = 40$ as in the examples for polynomial regression, then for $q = 2$, the design space consists of $40^2 = 1600$ points and for $q = 3$, the number of points increases to $40^3 = 64,000$. The computational complexity would thus increase with the dimensionality and thus the algorithm of §2.3 may not produce the optimal design in any reasonable amount of time. However, since $\mathcal{S}$ is symmetric about $\mathbf{0}$ and invariant under permutations of the coordinate axes, Fang and Wiens (2000) proposed that the designs in $\mathcal{S}$ can be generated by symmetrically choosing n_0 points in the x_1-axis and then forming the $q-$fold Cartesian product of these points with themselves, whence $n = n_0^q$. The algorithm of §2.3 then becomes applicable in choosing the n_0 points. Here, we propose to adopt an alternative approach, previously employed by Heo, Schmuland and Wiens (2001). In this approach we continue to assume symmetry of designs about $\mathbf{0}$ and exchangeability of the coordinate axes. (We can often arrange symmetry through an affine transformation of the independent variables, in which case we do not lose generality.) Invariance under permutations of the axes would be reasonable when there is no *a priori* reason to prefer one coordinate over another. To describe the approach, suppose the problem of interest requires n design points. We choose $n_0 = n/8$ points (x_1, x_2) in $\{0 \leq x_2 \leq x_1 \leq 1/2\}$ and then obtained the remaining $7n_0$ points by symmetry and exchangeability. Given the finite design space with $N = N_0^2$ points generated by the $2-$fold Cartesian product of equally spaced N_0 points in $\left[-\frac{1}{2}, \frac{1}{2}\right]$, this requires a search over the $N_0(N_0+1)/8$ points satisfying $\{0 \le x_2 \le x_1 \le 1/2\}$. We again employ simulated annealing.

Example 3. Partial second order model with two interacting regressors. We consider the model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + f(\mathbf{x}) + \varepsilon$$

on $\mathcal{S} = \left[-\frac{1}{2}, \frac{1}{2}\right] \times \left[-\frac{1}{2}, \frac{1}{2}\right]$, $\mathbf{x} = (x_1, x_2)^T$ with two interacting regressors, that is $z(\mathbf{x}) = (1, x_1, x_2, x_1 x_2)^T$. This model is exchangeable in x_1 and x_2. For $N_0 = 20$ and $n_0 = 40$, we obtain designs corresponding to $\rho = 1$ (minimizing variance only), .75, .5 and 0 (minimizing bias only). The variance minimizing design has five support points. The number of support points increases with increasing bias in the model. In general, the pattern of results agree with previous work in the literature. The designs are clusters of observations about the points that would have been the design points for the variance-minimizing design and the spread of the clusters increases with increasing bias. (see Figure 7 and Table 2).

Example 4. Full second-order model with $q = 2$. Here we consider the model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{12} x_1 x_2 + \beta_{22} x_2^2 + f(\mathbf{x}) + \varepsilon$$

on $\mathcal{S} = \left[-\frac{1}{2}, \frac{1}{2}\right] \times \left[-\frac{1}{2}, \frac{1}{2}\right]$, $\mathbf{x} = (x_1, x_2)^T$ and $z(\mathbf{x}) = (1, x_1, x_2, x_1 x_2, x_1^2, x_2^2)^T$. This model is an extension of that considered in the last example. It is also exchangeable in x_1 and x_2. Given $N_0 = 20$ and $n_0 = 55$, we obtained designs corresponding to $\rho = 1$ (minimizing variance only), .75, .5 and 0 (minimizing bias only). The pattern of results is again similar to those in the previous example. The number of distinct design points for each value of ρ is greater than that for the corresponding design for partial second-order model since we have more parameters here. See the design plots in Figure 8.

2.6 Concluding remarks

We have investigated new criteria for the construction of robust integer-valued regression designs. These designs are robust against misspecifications in the assumed model form. We recommend the use of the minave and robust prediction constrained criteria when the focus of the design is prediction while the robust estimation criterion is recommended when the design interest is estimation. Simulated annealing has been used successfully to seek integer designs both for unconstrained and constrained optimizations. The results obtained from the use of the three criteria follow a consistent pattern.

The patterns of the results are still similar to those obtained using a minimax approach even though the rationale for the designs presented here are based on the notion of averaging, rather than maximizing, the loss over the contamination space. The key message remains that the designs that protect against the general forms of model misspecification may be approximated by taking clusters of observations about those points that would have otherwise serve as the design points for variance-minimizing designs. We have also demonstrated the superiority of an integer-valued construction over the approximate continuous method in that as the desired number of observations increase our design does not merely produce replicates at the points that have been included in the design when fewer observations were taken; it instead incorporates additional design points.

Acknowledgements

The research of both authors is supported by the Natural Sciences and Engineering Research Council of Canada. We appreciate the suggestions of an anonymous referee.

References

Adewale, A.J., and Wiens, D.P. (2005), "Robust Designs for Misspecified Logistic Models," University of Alberta Statistics Centre Technical Report 05-08.

Bohachevsky, I.O., Johnson, M.E., and Stein, M.L. (1986), "Generalized Simulated Annealing for Function Optimization," *Technometrics*, 28, 209-217.

Box, G.E.P., and Draper, N.R. (1959), "A basis for the selection of a response surface design," *Journal of the American Statistical Association*, 54, 622 - 654.

Chen, R.-B. and Huang, M.-N. L. (2000), "Exact D-optimal Designs for Weighted Polynomial Regression Model," *Computational Statistics and Data Analysis*, 33, 137-149.

Fang, Z. and Wiens, D.P. (2000), "Integer-Valued, Minimax Robust Designs for Estimation and Extrapolation in Heteroscedastic, Approximately Linear Models," *Journal of the American Statistical Association*, 95 (451), 807-818.

Gaffke, N. (1987), "On D-optimality of Exact Linear Regression Designs with Minimum Support," *Journal of Statistical Planning and Inference*, 15, 189-204.

Guest, P.G. (1958), "The Spacing of Observations in Polynomial Regression," *The Annals of Mathematical Statistics*, 29 (1), 294-299.

Haines, L.M. (1987), "The Application of the Annealing Algorithm to the Construction of Exact Optimal Designs for Linear-Regression Models," *Technometrics*, 29, 439-447.

Heo, G., Schmuland, B., and Wiens, D.P. (2001), "Restricted Minimax Robust Designs for Misspecified Regression Models," *The Canadian Journal of Statistics*, 29, 117-128.

Hoel, P.G. (1958), "Efficiency Problems in Polynomial Estimation," *The Annals of Mathematical Statistics*, 29 (4), 1134-1145.

Huang, M.-N. L. (1987), "Exact D-Optimal Design for Polynomial Regression," *Bulletin of the Institute of Mathematics Academia Sinica*, 15, 59-71.

Läuter, E. (1974), "Experimental Design in a Class of Models," *Mathematische Operationsforschung und Statistik* 5, 379-396.

Läuter, E. (1976), "Optimal Multipurpose Designs for Regression Models," *Mathematische Operationsforschung und Statistik* 7, 51-68.

Montepiedra and Fedorov (1997), "Minimum Bias Designs with Constraints," *Journal of Statistical Planning and Inference*, 63, 97-111.

Pukelsheim, F. (1993), *Optimal Design of Experiments*, New York: Wiley.

CHAPTER III

ROBUST DESIGNS FOR MISSPECIFIED LOGISTIC MODELS

Abstract We develop criteria that generate robust designs and use such criteria for the construction of designs that insure against possible misspecifications in logistic regression models. The design criteria we propose are different from the classical in that we do not focus on sampling error alone. Instead we use design criteria that account as well for error due to bias engendered by the model misspecification.

Our robust designs optimize the average of a function of the sampling error and bias error over a specified misspecification neighbourhood. Examples of robust designs for logistic models are presented, including a case-study implementing the methodologies using beetle mortality data.

3.1 Introduction

Experimental designs have been treated extensively in the statistical literature, starting with designs for linear models and extending to non-linear models. A large volume of literature is devoted to designs assuming the exact correctness of the relationship between the response variable and the design (explanatory) variables. Box and Draper (1959) added another dimension to the theory by investigating the impact of model

---

[2] Co-authored with Professor Douglas P. Wiens. Submitted for publication.

misspecification. Following the work of Box and Draper the literature has since been replete with regression designs which are robust against violations of various model assumptions - linearity of the response, independence and homoscedasticity of the errors, etc. Authors who have considered designs with an eye on the approximate nature of the assumed linear models include Marcus and Sacks (1976), Li and Notz (1982), Wiens (1992), Wiens and Zhou (1999), to mention but a few.

For nonlinear designs, Fedorov (1972), Ford and Silvey (1980), Chaloner and Larntz (1989) and Chaudhuri and Mykland (1993) have explored the construction of optimal designs while assuming that the nonlinear model of interest is correctly specified. Still others have investigated designs for generalized linear models, a class of possibly nonlinear models in which the response follows a distribution from the exponential family such as normal, binomial, Poisson or gamma (McCullagh and Nelder 1989). The expository article Ford, Titterington and Kitsos (1989) hinted that in the context of nonlinear models, as in the case of linear model, the misspecification of the model itself is of serious concern. They asserted that "indeed, if the model is seriously in doubt, the forms of design that we have considered may be completely inappropriate." Sinha and Wiens (2002) have explored some designs for nonlinear models with due consideration for the approximate nature of the assumed model. In this work we consider designs for misspecified logistic regression models.

For the logistic model, the mean response $E\left(Y\right) = \mu$ depends on the parameters, $\boldsymbol{\beta}$, and the vector of explanatory variables, $\mathbf{x}$, through the nonlinear function $\mu = e^{\eta}/\left(1 + e^{\eta}\right)$, where $\eta = \mathbf{z}^{T}\left(\mathbf{x}\right)\boldsymbol{\beta}$. The function η is termed the linear predictor, with regressors $z_1\left(\mathbf{x}\right), ..., z_p\left(\mathbf{x}\right)$ depending on the q-dimensional independent variable $\mathbf{x}$.

The variance of the response, written $var\,(Y|\mathbf{x})$, is a nonlinear function of the linear predictor. Abdelbasit and Plackett (1983), Minkin (1987), Ford, Torsney and Wu (1992), Chaudhuri and Mykland (1993), Burridge and Sebastiani (1994), Atkinson and Haines (1996) and King and Wong (2000) have investigated designs for binary data, and in particular for logistic regression. As illustrated in these papers, the general approach to optimal design is to seek a design that optimizes certain functions of the information matrix of the model parameters. The information matrix for $\boldsymbol{\beta}$ from a design consisting of the points $\mathbf{x}_1 ..., \mathbf{x}_n$ is given by

$$\sum_{i=1}^{n} w(\mathbf{x}_{i,}\ \boldsymbol{\beta})\mathbf{z}\left(\mathbf{x}_i\right)\mathbf{z}^T\left(\mathbf{x}_i\right) = \mathbf{Z}^T\mathbf{W}\mathbf{Z}, \tag{3.1}$$

where $\mathbf{Z} = \left(\mathbf{z}\left(\mathbf{x}_1\right), \mathbf{z}\left(\mathbf{x}_2\right), ..., \mathbf{z}\left(\mathbf{x}_n\right)\right)^T$ and $\mathbf{W} = diag\left(w(\mathbf{x}_{1,}\boldsymbol{\beta}), w(\mathbf{x}_{2,}\boldsymbol{\beta}), ..., w(\mathbf{x}_{n,}\boldsymbol{\beta})\right)$ for weights $w(\mathbf{x}_{i,}\ \boldsymbol{\beta}) = \left(d\mu/d\eta_i\right)^2 / var\,(Y|\mathbf{x}_i)$. Thus, as with nonlinear experiments the information matrix depends on the unknown parameters $\boldsymbol{\beta}$. Designing an experiment for the estimation of these parameters would then seem to require that these parameters be known! The following are some of the approaches that have been explored in the literature for handling the dependency of the information matrix on $\boldsymbol{\beta}$.

1. Locally optimal designs: A traditional approach in designing a nonlinear experiment is to aim for maximum efficiency at a best guess (initial estimate) of the parameter values (Chernoff 1953). Designs that are optimal for given parameter values are dubbed locally optimal designs. These designs may be stable over a range of parameter values. However, if unstable, a design which is optimal for a best guess may not be efficient for parameter values in even a small neighbourhood of this guess.

2. Bayesian optimal designs: A natural generalization of locally optimal designs is to use a prior distribution on the unknown parameters rather than a single guess. The approach which assumes such a prior and incorporates this distribution into the appropriate design criteria is termed Bayesian optimal design - see Chaloner and Larntz (1989) and Dette and Wong (1996).

3. Minimax optimal designs: Rather than assume a prior distribution, this approach assumes a range of plausible values for the parameters. The minimax optimal design is the design with the least loss when the parameters take the worst possible value within their respective ranges. These least favourable parameter values are those that maximize the loss (King and Wong 2000; Dette, Haines and Imhof 2003).

4. Sequential designs: In sequential designs, the experiment is done in stages. Parameter estimates from a previous stage are used as initial estimates in the current stage. The process continues until optimal designs are obtained (Abdelbasit and Plackett 1983; Sinha and Wiens 2002).

Suppose an experimenter is faced with a set $\mathcal{S} = \{\mathbf{x}_i\}_{i=1}^{N}$ of possible design points from which he is interested in choosing n, not necessarily distinct, points at which to observe a binary response Y. The experimenter makes $n_i \geq 0$ observations at $\mathbf{x}_i$ such that $\sum_{i=1}^{N} n_i = n$. The design problem is to choose $n_1, ..., n_N$ in an optimal manner. The objective then is to choose a probability distribution $\{p_i\}_{i=1}^{N}$, with $p_i = n_i/n$, on the design space $\mathcal{S}$. The commonalities in the work of the authors who have considered logistic design is the salient assumption that the assumed model form is exactly correct. In this work, we propose the construction of robust designs for logistic models with due consideration for possible misspecification in the assumed form of the systematic component - the linear predictor. The linear predictor could be said to be misspecified when it does not reflect the influence of the covariates correctly, possibly due to omitted covariates or to omission of some transformation of existing covariates in the model. In this section we formalize our notion of model misspecification.

We suppose that the experimenter fits a logistic model with the mean response

$$\mu_i = \mu(\eta_i), \quad i = 1, ..., n, \tag{3.2}$$

for $\eta_i = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta}_0$ when in fact the true mean response is represented by

$$\mu_{T,i} = \mu(\eta_i + f(\mathbf{x}_i)). \tag{3.3}$$

The target parameter $\boldsymbol{\beta}_0$ is defined by

$$\boldsymbol{\beta}_0 = \arg\min_{\beta} \frac{1}{N} \sum_{i=1}^{N} \left\{ E(Y|\mathbf{x}_i) - \mu\left(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta}\right) \right\}^2.$$

Thus the target parameter is that which guarantees the least sum of squares of discrepancies, over all points in the design space, between the assumed mean response and the true mean response. The contamination function $f(\mathbf{x})$ may or may not be known. It would be known, for example, when an experimenter decides to fit the more parsimonious model (3.2) despite the knowledge of a more appropriate model (3.3) with a specified $f(\mathbf{x})$. For instance, the simplified model might be required if the number of support points is not sufficient to handle a more complicated but more appropriate model. Knowing that the parsimonious model might result in an inferior analysis, the experimenter may seek a design that remedies the anticipated model inadequacy.

The contamination function would be unknown in a situation where the experimenter is aware of the possible uncertainties in the assumed model form and might have clues about the properties of the possible misspecification, but not know its exact structure. When $f(\mathbf{x})$ is unknown, some knowledge about its properties or conditions it satisfies would be required to construct any appropriate design. This is so because no single design which takes a finite number of observations can protect against all possible forms of bias. Thus, we must impose some conditions on the contamination function when its precise form is unknown.

To bound the bias of an estimator $\hat{\beta}$, we assume that

$$\frac{1}{N}\sum_{i=1}^{N} f^2(\mathbf{x}_i) \leq \tau^2 \tag{3.4}$$

with $\tau^2 = O(n^{-1})$. This latter requirement is analogous to the notion of contiguity in the asymptotic theory of hypothesis testing, and is justified in the same manner. The choice of τ is discussed following Theorem 3.2 in the next section. In order to ensure identifiability of the model parameters $\boldsymbol{\beta}$ and the contamination function $f(\mathbf{x})$ we require that the vector of regressors and the contamination be orthogonal. That is,

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}(\mathbf{x}_i) f(\mathbf{x}_i) = \mathbf{0}. \tag{3.5}$$

Let $\mathcal{F}$ denote the class of contamination functions $f(\mathbf{x})$ satisfying (3.4) and (3.5).

3.2 Loss Functions: Estimated and Averaged Mean Squared Errors of Prediction

The basis for the construction of classical designs for logistic regression models has typically been the minimization of (a function of) the inverse of Fisher's information matrix (3.1) - see Atkinson and Haines (1996). However, in the face of model misspecification the asymptotic covariance, $cov\left(\hat{\boldsymbol{\beta}}\right)$, of the maximum likelihood estimator of the model parameters no longer equals the inverse of Fisher information - see White (1982) and also Fahrmeir (1990), who discusses the asymptotic properties of MLEs under a misspecified likelihood.

Suppose that data $\{\mathbf{x}_i, y_i\}$ are given, where the $\mathbf{x}_i$ are the design points chosen from $\mathcal{S}$ with n_i observations at $\mathbf{x}_i$ such that $\sum_{i=1}^{N} n_i = n$, and y_i is the proportion of successes at location $\mathbf{x}_i$. The asymptotic bias and covariance of the MLE $\hat{\boldsymbol{\beta}}$ are given in Theorem 3.1 below; see the Appendix for details of this and other proofs. The expressions for the asymptotic bias and covariance of the MLE $\hat{\boldsymbol{\beta}}$ is used in the derivation of the loss function in Corollary 3.1.

Theorem 3.1 *Define*

$$w_i = \frac{d\mu_i}{d\eta_i} = \mu_i\left(1-\mu_i\right) = \frac{1}{4}sech^2\left(\frac{\mathbf{z}^T\left(\mathbf{x}_i\right)\boldsymbol{\beta}_0}{2}\right), \tag{3.6}$$

and let $\mathbf{Z}$ *be the* $N \times p$ *matrix with rows* $\mathbf{z}^T\left(\mathbf{x}_i\right)$. *Recall (3.2) and (3.3); let* $\boldsymbol{\gamma}$ *and* $\boldsymbol{\gamma}_T$ *be the* $N \times 1$ *vectors with elements* μ_i *and* $\mu_{T,i}$ *respectively. Let* $\mathbf{P}$, $\mathbf{W}$ *and* $\mathbf{W}_T$ *be the* $N \times N$ *diagonal matrices with diagonal elements* n_i/n, w_i *and* $w_{T,i} = \mu_{T,i}\left(1-\mu_{T,i}\right)$ *respectively. Finally, define* $\mathbf{b} = \mathbf{Z}^T\mathbf{P}\left(\boldsymbol{\gamma}_T - \boldsymbol{\gamma}\right)$, $\mathbf{H}_n = \mathbf{Z}^T\mathbf{PWZ}$, $\tilde{\mathbf{H}}_n = \mathbf{Z}^T\mathbf{PW}_T\mathbf{Z}$. *The asymptotic bias and asymptotic covariance matrix of the maximum likelihood* *estimator $\hat{\boldsymbol{\beta}}$ of the model parameter vector $\boldsymbol{\beta}$ from the misspecified model are*

$$bias\left(\hat{\boldsymbol{\beta}}\right) = E\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) = \mathbf{H}_n^{-1}\mathbf{b} + o\left(n^{-1/2}\right),$$

$$cov\left(\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)\right) = \mathbf{H}_n^{-1}\tilde{\mathbf{H}}_n\mathbf{H}_n^{-1} + o\left(1\right),$$

respectively.

Since the typical focus of logistic designs is prediction, we take as loss function the normalized average mean squared error (AMSE) I of the response prediction $\mu\left(\hat{\eta}_i\right)$, with $\hat{\eta}_i = \mathbf{z}^T\left(\mathbf{x}_i\right)\hat{\boldsymbol{\beta}}$. This is given by

$$I \triangleq \frac{n}{N}\sum_{i=1}^{N} E\left[\left\{\mu\left(\hat{\eta}_i\right) - \mu\left(\eta_i + f\left(\mathbf{x}_i\right)\right)\right\}^2\right].$$

Corollary 3.1 *The AMSE has the asymptotic approximation $I = \mathcal{L}_I\left(\mathbf{P}, \mathbf{f}\right) + o(1)$, where*

$$\mathcal{L}_I\left(\mathbf{P}, \mathbf{f}\right) = \frac{1}{N}\left\{tr\left[\mathbf{W}\mathbf{Z}\mathbf{H}_n^{-1}\tilde{\mathbf{H}}_n\mathbf{H}_n^{-1}\mathbf{Z}^T\mathbf{W}\right] + n\left\|\mathbf{W}\left(\mathbf{Z}\mathbf{H}_n^{-1}\mathbf{b} - \mathbf{f}\right)\right\|^2\right\} \tag{3.7}$$

for $\mathbf{f} = \left(f\left(\mathbf{x}_1\right), ..., f\left(\mathbf{x}_N\right)\right)^T$.

By using the expressions for asymptotic bias and covariance given in Theorem 3.1, Corollary 3.1 expresses the average mean squared error (AMSE) as an explicit function of the design matrix $\mathbf{Z}$ and contamination vector $\mathbf{f}$. The first term in the loss function $\mathcal{L}_I$ corresponds to the average variance of the predictions and it depends on the contamination function $f\left(\mathbf{x}\right)$ through the matrix $\tilde{\mathbf{H}}_n$. The second term in the expression for $\mathcal{L}_I$ is the average squared bias of the predictions, which depends on the contamination $f\left(\mathbf{x}\right)$ through the contamination vector $\mathbf{f}$ and implicitly through the vector **b**. Thus a design cannot minimize (3.7) directly without certain assumptions about the contamination $f(\mathbf{x})$.

Fang and Wiens (2000) constructed integer-valued designs for linear models, in the case of an unknown f, using a minimax approach. Their minimax criterion minimizes the maximum value of the loss function over f. They solve the design problem by minimizing the loss when the misspecification is the worst possible in the neighbourhood of interest.

Here, we take one of two approaches depending on whether or not there are initial data. If we have initial data we represent the discrepancy between the true response and the assumed response, at a sampled location **x**, by

$$d(\mathbf{x}) = \mu\left(\mathbf{z}^T(\mathbf{x})\boldsymbol{\beta}_0 + f(\mathbf{x})\right) - \mu\left(\mathbf{z}^T(\mathbf{x})\boldsymbol{\beta}_0\right),$$

and estimate this by the residual $\hat{d}(\mathbf{x}) = y(\mathbf{x}) - \mu\left(\mathbf{z}^T(\mathbf{x})\hat{\boldsymbol{\beta}}\right)$. A first order approximation is $d(\mathbf{x}) \approx (d\mu/d\eta) f(\mathbf{x})$, leading to $\hat{f}(\mathbf{x}) = \hat{d}(\mathbf{x}) / \left(d\mu/d\eta_{|\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}\right)$. We smooth this estimated contamination over the entire design space - see Example 3 of §3.3.2 for an illustration. The resulting estimate $\hat{\mathbf{f}}$, together with $\hat{\boldsymbol{\beta}}$, is then substituted into the terms in (3.7), and we compute a design minimizing $\mathcal{L}_I\left(\mathbf{P}, \hat{\mathbf{f}}\right)$ using the techniques outlined in §3.3.1.

If there are no initial data we propose to instead average $\mathcal{L}_I$ over $\mathcal{F}$ defined by (3.4) and (3.5). Our optimal design minimizes this average value. This criterion is in the spirit of Läuter (1974, 1976). Läuter's criterion optimizes the weighted average of the loss of a finite set of plausible models. Here we are instead faced with an infinite set of models indexed by $f \in \mathcal{F}$.

To carry out the averaging we begin as in Fang and Wiens (2000), with the singular value decomposition

$$\mathbf{Z} = \mathbf{U}_{N\times p}\mathbf{\Lambda}_{p\times p}\mathbf{V}_{p\times p}^{T}, \tag{3.8}$$

with $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}_p$ and $\mathbf{\Lambda}$ diagonal and invertible. We augment $\mathbf{U}$ by $\tilde{\mathbf{U}}_{N\times(N-p)}$ such that $\left[\mathbf{U} \vdots \tilde{\mathbf{U}}\right]_{N\times N}$ is orthogonal. Then by (3.4) and (3.5), we have that there is an $(N-p)\times 1$ vector $\mathbf{t}$, with $\|\mathbf{t}\| \leq 1$, satisfying

$$\mathbf{f}\ (= \mathbf{f}_\mathbf{t}) = \tau\sqrt{N}\tilde{\mathbf{U}}\mathbf{t}. \tag{3.9}$$

The average loss is taken to be the expected value of (3.7), as a function of $\mathbf{t}$, with respect to the uniform measure on the unit sphere and its interior in $\mathbb{R}^{N-p}$. This measure has density $p(\mathbf{t}) = \frac{1}{\kappa_{N,p}} I\left(\|\mathbf{t}\| \leq 1\right)$, where $\kappa_{N,p} = \pi^{(N-p)/2}/\Gamma\left(\frac{N-p}{2}+1\right)$ is the volume of the unit sphere. Theorem 3.2 handles the averaging of $\mathcal{L}_I$. The importance of this theorem is in its elimination of the dependency of our design criterion on the unknown contamination function.

Theorem 3.2 *The average loss over the misspecification neighbourhood $\mathcal{F}$ is, apart from terms which are $o(1)$, given by*

$$\begin{aligned}\mathcal{L}_{I,ave}(\mathbf{P},\rho) &\triangleq \int \mathcal{L}_I(\mathbf{P},\mathbf{f}_\mathbf{t})\,p(\mathbf{t})\,d\mathbf{t} \\ &= \frac{1}{N} tr\left[\left(\mathbf{U}^T\mathbf{PWU}\right)^{-1}\left(\mathbf{U}^T\mathbf{W}^2\mathbf{U}\right)\right] + \frac{\rho}{N-p+2} tr\left[\mathbf{W}\left(\mathbf{R}-\mathbf{I}_N\right)\left(\mathbf{R}^T-\mathbf{I}_N\right)\mathbf{W}\right],\end{aligned} \tag{3.10}$$

where $\rho = n\tau^2$ and $\mathbf{R}_{N\times N} = \mathbf{U}\left(\mathbf{U}^T\mathbf{PWU}\right)^{-1}\mathbf{U}^T\mathbf{PW}$. For numerical work it is more efficient to compute the second trace in (3.10) as

$$tr\left[\mathbf{W}\left(\mathbf{R}-\mathbf{I}_N\right)\left(\mathbf{R}^T-\mathbf{I}_N\right)\mathbf{W}\right] = \sum_{i=1}^{N} w_i^2\left\|\tilde{\mathbf{r}}_i\right\|^2,$$

where $\tilde{\mathbf{r}}_i^T$ *is the* i^{th} *row of* $\mathbf{R} - \mathbf{I}_N$.

The dependency of the design criterion on the unknown contamination has now been represented by a design parameter ρ, which can be chosen by the experimenter. This parameter can be interpreted as a measure of departure of the true model from the fitted model. In other words, it is a measure of the experimenter's lack of confidence in the validity of the model that he fits. If he believes that this assumed model is exactly correct, he chooses $\rho = 0$ corresponding to the classical I-optimal design. On the other hand, if the experimenter believes that the assumed model is highly uncertain, he chooses a large value of ρ for his design. Designs corresponding to a large value of ρ are dominated by the bias component of the loss.

Our design criterion (3.10) remains dependent on the model parameter vector $\boldsymbol{\beta}_0$, as is the case in the general nonlinear design problems, through the weights, as at (3.6). In the examples of the next section we handle this dependency by either taking a guess (locally optimal designs) or assuming a prior distribution, say $\pi(\boldsymbol{\beta}_0)$, on $\boldsymbol{\beta}_0$ (Bayesian designs). The loss function $\mathcal{L}_{I,ave}$ is modified as $\int \mathcal{L}_{I,ave}(\boldsymbol{\beta})\,\pi(\boldsymbol{\beta})\,d\boldsymbol{\beta}$ in the case of a Bayesian construction.

3.3 Designs - Algorithm and Examples

3.3.1 Simulated Annealing

We consider problems with polynomial predictors, viz. $\eta = \mathbf{z}^T(x)\boldsymbol{\beta}$ with $\mathbf{z}(x) = (1, x, x^2, ..., x^{p-1})^T$. We take equally spaced design points $\{x_i\}_{i=1}^N$ in the interval $\mathcal{S}$. Our design minimizes the relevant loss function through the matrix

$$\mathbf{P} = diag(n_1/n, \cdots, n_N/n).$$

This is a nonlinear integer optimization problem for which there is no analytic solution, and for which we employ simulated annealing to search for the optimal design.

The simulated annealing algorithm is a direct search random walk optimization algorithm which has been quite successful at finding global extrema of non-smooth functions and/or functions with many local extrema. The algorithm consists of three steps, each of which must be well adapted to the problem of interest for the algorithm to be successful. The first step is a specification of the initial state of the process. In this step an initial design has to be specified, say $\mathbf{P}_0$. The second is a specification of a scheme by which a new design $\mathbf{P}_1$ is chosen from the optimization space. The last step is a prescription of the basis of acceptance or rejection: an acceptance with probability 1 if $\mathcal{L}_{I,ave}(\mathbf{P}_1) < \mathcal{L}_{I,ave}(\mathbf{P}_0)$, otherwise acceptance with probability $\exp\{-(\mathcal{L}_{I,ave}(\mathbf{P}_1) - \mathcal{L}_{I,ave}(\mathbf{P}_0))/T\}$, where T is a tuning parameter. The tuning parameter is usually decreased as the iterations proceed. After a large number of iterations between the second and third steps the loss function is expected to converge to its (near) minimum value. Simulated annealing has been used for design problems by, among others, Meyer and Nachtsheim (1988), Fang and Wiens (2000) and Adewale and Wiens (2006).

A very simple and general approach that we considered for choosing the initial design is to randomly select p points from $\{x_i\}_{i=1}^N$ and randomly allocate the observations to these points such that the total number of observations is n. Fang and Wiens (2000) used a different approach which assumes that one of (n, N) is a multiple of the other. For any (n, N) combination they chose the initial design to be as uniform as possible. We applied this approach as well but found that the two approaches are

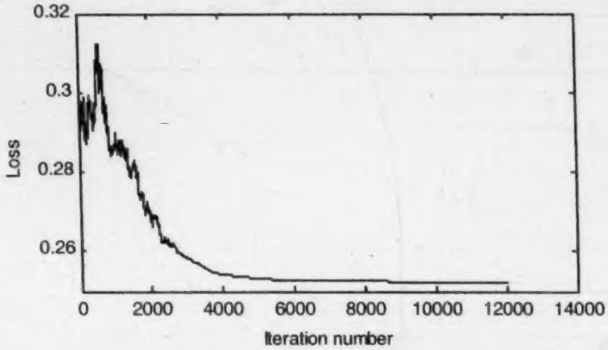

Figure 9: Simulated Annealing Trajectory for Logistic Design with $\eta = 1 + 3x$, $x \in \mathcal{S} = [-1, 1]$, $\rho = 0$ and $(N = 40, n = 200)$.

equally efficient. For generating a new design we adopted the perturbation scheme of Fang and Wiens (2000). The turning parameter in the third step was chosen initially such that the acceptance rate is in the range 70% and 95%. We decrease T by a factor of .95 after each 20 iterations. In the examples below we run the algorithm several times with varying turning parameter specification and reduction rate in order to satisfy ourselves that the resulting design has the least loss possible under the relevant circumstances of each example. In Figure 9 we present the simulated annealing trajectory for one of the cases presented in Example 1. It took 82.69 seconds for the algorithm to complete the preset maximum number of iterations (12000, for this case) and the minimum loss was attained just before the 9000^{th} iteration.

3.3.2 Examples

Example 1: No contamination. As a benchmark we first consider the logistic regression model with a single predictor: $p = 2$, $\mathbf{z}(x) = (1, x)^T$, $x \in \mathcal{S} = [-1, 1]$, $\boldsymbol{\beta} =$

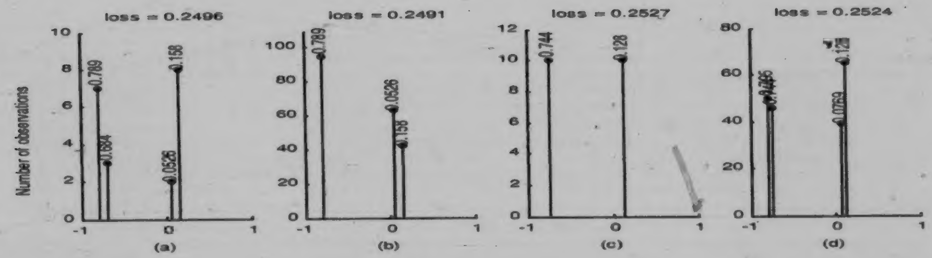

Figure 10: Locally optimal designs minimizing $\mathcal{L}_{I,ave}$ when $\rho = 0$ (no contamination) with (a) $N = 20, n = 20$; (b) $N = 20, n = 200$; (c) $N = 40, n = 20$; (d) $N = 40, n = 200$.

$(1,3)^T$, and no contamination: $\rho = 0$. We initially took $n = 20$, $N = 40$ and considered designs minimizing $\mathcal{L}_I$. The annealing algorithm converged to the design placing 10 of the 20 observations at each of the points $-.744$ and $.128$. This design is therefore the classical $I-$optimal design minimizing the averaged variance of the predictions over $\mathcal{S}$. There is evidently no previous theory that applies to this case. However, using a model that is a reparameterization of ours, and a continuous design space $[-1, 1]$, King and Wong (2000) showed the locally $D-$optimal design to be the design that is equally supported at $-.848$ and $.181$. For the sake of comparison, we sought an equivalent design using our finite design space and the algorithm described above. The resulting design places 10 of 20 observations at each of $-.846$ and $.180$. Thus, our algorithm attains the closest approximation to King and Wong's solution in that the points $-.846$ and $.180$ are nearest, in our design space, to $-.848$ and $.181$. Unlike designs for linear models, the optimal designs in this case do not necessarily place observations at the extreme points of the design space. This phenomenon is due to the curvature introduced by the link function and the resulting nonlinear relationship between the mean response and x.

Table 3: Comparing Restricted Designs with Unrestricted Design; $\rho = 0$.

(N, n)	Restricted Design (Two-point)		Unrestricted Design	
	Design Points[1]	Loss	Design Points[1]	Loss
$(20, 20)$	$-.789(9)$, $.053(11)$	.250	$-.789(7)$, $-.684(3)$, $.0526(2)$, $.158(8)$	.250
$(20, 200)$	$-.789(94)$, $.053(106)$	.250	$-.789(95)$, $.0526(63)$, $.158(42)$	.249
$(40, 20)$	$-.744(10)$, $.128(10)$	.2527	$-.744(10)$, $.128(10)$	.2527
$(40, 200)$	$-.744(97)$, $.128(103)$	.2525	$-.795(49)$, $-.744(47)$, $.0769(39)$, $.128(65)$	.2524

[1] Number of observations in parentheses

Our numerical results further revealed that the designs depend on the number of points in the design space and the number of observations the experimenter want to take. For this "no-contamination" case, we investigated designs for various combinations of N and n. Some of these designs are presented in Figure 10. The number of distinct design points varies from 2 to 4. We found this somewhat surprising, in light of the fact that all $D-$optimum designs for the two parameter logistic model in the literature are two-point designs. Presumably this is explained through our use of a finite design space, and/or our use of average loss rather than that based on the determinant of the information matrix.

To check that this phenomenon was not merely an artifact due to a lack of convergence, we modified our algorithm to obtain 'restricted' designs - restricted to two support points only. The results for the same values of N and n as in Figure 10 are presented in Table 3. The loss for the unrestricted design is less than or equal to that for the corresponding restricted design in all cases considered.

In the examples that follow we limit discussion to the case $N = 40$, $n = 200$.

Example 2: Example 1 continued. In this example, which we include largely for illustrative purposes, the form of the contamination is known. Suppose that the experimenter anticipates fitting a simple logistic model, while wishing protection against a range of logistic models with quadratic predictor: $\eta(x) = \mathbf{z}^T(x)\boldsymbol{\beta} + f(x)$, where $\mathbf{z}^T(x)$ and $\boldsymbol{\beta}$ are as in Example 1, and $f(x) = \beta_2(x^2 - \mu_2)/\sqrt{\mu_4 - \mu_2^2}$, for $\mu_k = N^{-1}\sum x_i^k$ ($= 0$ if k is odd). The contaminant $f(x)$ is an omitted quadratic term, translated and scaled to ensure the orthogonality condition (3.5); (3.4) becomes $|\beta_2| \leq \tau$. We obtained optimal designs for various values of the quadratic coefficient β_2. The resulting designs and corresponding values of the loss function are presented in Table 4. In the range of values of β_2 considered, we found that the number of distinct points varied from 3 to 6. The spread of the design over the design space tended to increase as the magnitude of the omitted quadratic term increases. We computed the premium paid for robustness and the gain due to robustness for each design presented as

$$\textit{Premium} = \left(\frac{\mathcal{L}_I(\mathbf{P}_{opt}, \mathbf{f} = \mathbf{0})}{\mathcal{L}_I(\mathbf{P}_{classical}, \mathbf{f} = \mathbf{0})} - 1\right) \times 100\%, \tag{3.11}$$

and

$$\textit{Gain} = \left(1 - \frac{\mathcal{L}_I(\mathbf{P}_{opt}, \mathbf{f})}{\mathcal{L}_I(\mathbf{P}_{classical}, \mathbf{f})}\right) \times 100\%. \tag{3.12}$$

The gain measure is the percentage reduction in loss due to the use of a robust design as opposed to a (non-robust) classical design which assumes the fitted model to be exactly correct. The premium measure is the percentage increase in loss as a result of not using the classical design if in actual fact the assumed model is correct. The application of the premium and/or gain measure depends on the amount of confidence

Table 4: Designs for simple logistic model when the true model has a quadratic term.

β_2	Design Points (Number of Observations)	$\mathcal{L}_I(\mathbf{P}, \mathbf{f})$	Premium	Gain
−10	$-1\,(42)\,,\ -.180\,(42)\,,\ -.128\,(96)\,,$ $-.077\,(12)\,,\ -.026\,(2)\,,\ .077\,(6)$	3.500	35.0%	34.8%
−3	$-1\,(48)\,,\ -.282\,(26)\,,\ -.231\,(64)\,,\ .282\,(62)$	.5020	10.9%	17.0%
−1	$-.949\,(42)\,,\ -.590\,(34)\,,\ -.539\,(29)\,,\ .128\,(95)$	.2756	2.1%	.5%
0	$-.795\,(49)\,,\ -.744\,(47)\,,\ .077\,(39)\,,\ .128\,(65)$	.2524	0	0
1	$-.641\,(100)\,,\ .128\,(22)\,,\ .180\,(78)$	.3080	1.9%	4.0%
3	$-.692\,(57)\,,\ -.641\,(29)\,,\ -.077\,(39)\,,$ $-.026\,(44)\,,\ .795\,(31)$	.6073	11.9%	19.9%
10	$-1\,(11)\,,\ -.590\,(51)\,,\ -.539\,(27)\,,$ $-.231\,(30)\,,\ -.180\,(40)\,,\ .949\,(41)$	3.679	39.0%	42.9%

the experimenter has in his knowledge of the true model. In this example, since the assumption is that the experimenter knows that the model with a linear predictor involving the quadratic term is a more appropriate model, the relevant measure would be the gain. Nevertheless, both measures are reported in Table 4. The value of a design from our robust procedure increases with increasing magnitude of the quadratic parameter. On the other hand, the experimenter has to be aware of the increasing premium when his knowledge of the true model is not accurate. The premium paid for robustness also increases with the magnitude of the quadratic parameter.

Example 3: Designing when there are initial data to estimate contamination. Table 5 shows simulated data ("# of successes") from a logistic regression model with the predictor $\eta(x) = 1 + 3x + f(x)$, the model of the previous example; the quadratic parameter was $\beta_2 = 3$. The data were simulated using a uniform design over equally spaced points in $[-1, 1]$. Given the simulated data, we suppose the contamination function $f(x)$ to be unknown. We proceed using the procedure described in §3.2 for estimation and eventual smoothing of the contamination. A plot of the estimated

Table 5: Experimental design and response values.

Design point	-1	$-7/9$	$-5/9$	$-3/9$	$-1/9$	$1/9$	$3/9$	$5/9$	$7/9$	1
# of observations	20	20	20	20	20	20	20	20	20	20
# of successes	8	6	7	13	17	18	18	19	20	20

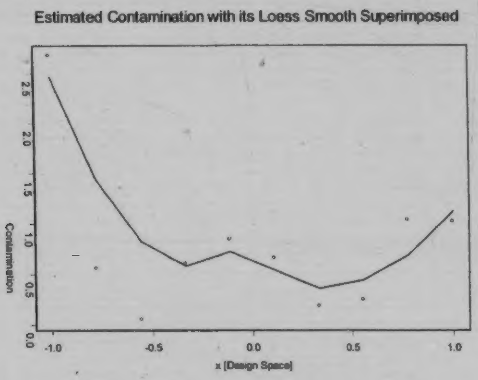

Figure 11: Estimated contamination plot for Example 3. True (but unknown) form of contamination is quadratic.

contamination with its loess smooth $\hat{f}(x)$ over the design space is presented in Figure 11.

We plugged the smoothed contamination values into the loss function (3.7), and used simulated annealing to obtain the design. Our design places 34, 82, and 84 of the 200 observations at $-.641$, $-.590$, and $.180$ respectively. For this design the premium for robustness is 5.0% and the gain is 60.0%. This example indicates that when there are initial data, it is expedient to incorporate the information from the data into the design procedure. The resulting design can lead to substantial gain at a reduced premium.

Example 4: Unknown contamination. Consider the logistic model with predictor

$$\eta(x) = \beta_0 + \beta_1 x + f(x). \tag{3.13}$$

In this example - as in Example 3 - we assume that f is an unknown member of the class $\mathcal{F}$ defined by (3.4) and (3.5). In Figure 12 we exhibit designs minimizing the averaged loss (3.10) for various values of ρ, β_0 and β_1. We observed a progression of the dispersion of the design points over the design space with increasing ρ. The pattern of the dispersion is, however, modified by the curvature indexed by β_0 and β_1 through the nonlinear mean response. For small ρ our robust designs can be described as taking clusters of observations at neighbouring locations rather than replications at only a few distinct sites; this was noticed for linear models by Fang and Wiens (2000). However, here there is always a pattern to the clusters of observation to be taken depending on the values of the model parameters. Large values of ρ connote large departures from the assumed model and an extremely large ρ value corresponds to the all-bias design. Even though the all-bias design is spread over the entire design space the frequencies of observations are different and these frequencies are prescribed by the curvature of the mean response as determined by the model parameters. In Table 6 we present the values of the premium paid and the gain in robustness for designs corresponding to different values of ρ for the particular case of $(\beta_0, \beta_1) = (1, 3)$. The gain in robustness, measured by (3.12), exceeds the premium paid, as measured by (3.11), for each design. Increasing robustness, however, comes with increasing premium; the experimenter would thus have to choose his level of comfort.

Thus far, the examples we have presented have been locally optimal, hence have assumed good parameter guesses for unknown model parameters. In the absence of

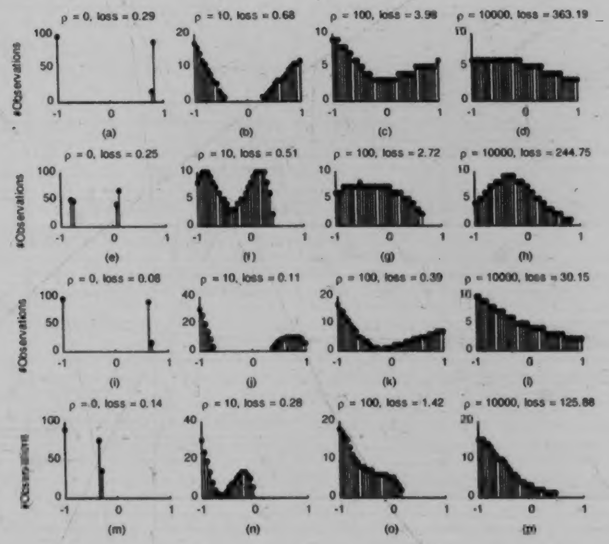

Figure 12: Locally optimal designs in Example 4; (a) - (d): $(\beta_0, \beta_1) = (1,1)$; (e)-(h): $(\beta_0, \beta_1) = (1,3)$; (i)-(l): $(\beta_0, \beta_1) = (3,1)$; (m)-(p): $(\beta_0, \beta_1) = (3,3)$.

a reliable best guess for model parameters, Sitter (1992) and King and Wong (2000) considered minimax D-optimality, a procedure which assumes the knowledge of a prior range for each of the parameters. We consider a Bayesian paradigm to be in the same spirit as averaging the contamination function over the specified misspecification neighbourhood, and take independent uniform prior distributions over the range of each model parameter. Our design criteria then becomes the expected loss, $E(\mathcal{L}_{I,ave}(\mathbf{P}, \rho))$, with the expectation taken with respect to the prior distribution. The dependency of our design criteria on the model parameters is through the weights w_i, and we do not have analytic expressions for the resulting integrals. In the examples that follow we employ number-theoretic methods for numerical evaluation of multiple integrals as discussed in Fang and Wang (1994). This approach is based

Table 6: Design for unknown contamination with $\beta_0 = 1$, $\beta_1 = 3$.

ρ	$\mathcal{L}_{I,ave}(\mathbf{P},\rho)$	Premium	Gain
0	.2524	0	0
1	.2809	.62%	2.01%
10	.5090	3.06%	14.33%
100	2.7204	7.44%	25.87%
1000	24.7294	11.29%	28.16%
10000	244.7545	11.97%	28.43%

on generating quasi-random points in the domain of definition of the integrand, and averaging the values of the loss over the sample of points.

Example 5: Robust Bayesian Design. In this example we consider the following ranges of parameter values: (a) $[.5, 1.5] \times [2.5, 3.5]$, (b) $[.5, 1.5] \times [1, 5]$, (c) $[-1, 3] \times [2.5, 3.5]$, (d) $[-1, 3] \times [1, 5]$, all with centre point $(1, 3)$ but with coverage areas 1, 4, 4, and 16 respectively. As described above, the robust design for each range of parameters values is the design that minimizes the expected average loss with respect to uniform distributions on the specified ranges of parameters values. For each of the designs - see Figure 13 - we take $\rho = .25$. We observed an increasing spread over the design space with increasing uncertainty in model parameters, as measured by the coverage area of the priors. This is consistent with previous work in optimal Bayesian design - see, for example, Chaloner and Larntz (1989) - which suggests increasing number of distinct design points with increasing uncertainty in the specified prior distributions. Comparing the design plots in panels (b) and (c) of Figure 13, we see that there is more sensitivity to uncertainty in the intercept parameter than the slope parameter.

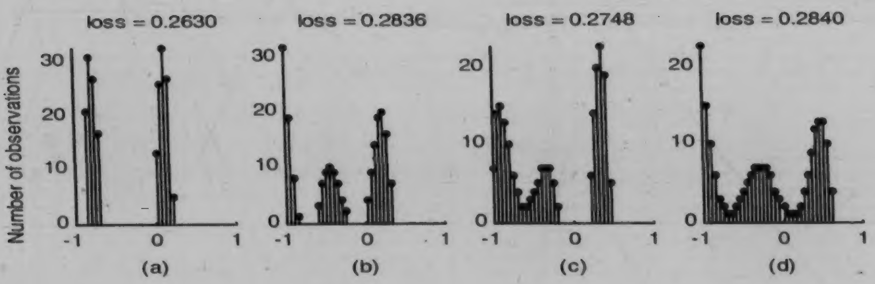

Figure 13: Robust Bayesian optimal design in Example 5 with $\rho = .25$ and parameters β_0 and β_1 having independent uniform priors over (a) $[.5, 1.5] \times [2.5, 3.5]$, (b) $[.5, 1.5] \times [1, 5]$, (c) $[-1, 3] \times [2.5, 3.5]$, (d) $[-1, 3] \times [1, 5]$.

Table 7: Beetle mortality data.

Dose, x_i ($\log_{10} CS_2 mgl^{-1}$)	1.69	1.72	1.75	1.78	1.81	1.84	1.86	1.88
Number of beetles, n_i	59	60	62	56	63	59	62	60
Number killed, $n_i y_i$	6	13	18	28	52	53	61	60

3.4 Case Study: Beetle Mortality Data

Bliss (1935) reported the numbers of beetles dead after five hours exposure to gaseous carbon disulphide at various concentrations. The doses are given in Table 7; to facilitate our discussion we have linearly transformed these to the range $[0, 1]$. Note that the original design is then very nearly uniform on the 8 equally spaced points $0\,(1/7)\,1$.

We first fitted the logistic model with the linear predictor $\eta^{(1)} = \beta_0^{(1)} + \beta_1^{(1)} x$, and obtained the estimates $\hat{\beta}_0^{(1)} = -2.777$ and $\hat{\beta}_1^{(1)} = 6.621$ with the estimated variance-covariance matrix

$$\Sigma^{(1)} = \begin{pmatrix} .082 & -.144 \\ -.144 & .317 \end{pmatrix}$$

and deviance $= 11.232$ ($df = 6$). The corresponding estimates for the logistic model with the linear predictor $\eta^{(2)} = \beta_0^{(2)} + \beta_1^{(2)} x + \beta_2^{(2)} x^2$ are $\hat{\beta}_0^{(2)} = -2.00$, $\hat{\beta}_1^{(2)} = 1.60$,

$\hat{\beta}_2^{(2)} = 5.84$ and

$$\Sigma^{(2)} = \begin{pmatrix} .124 & -.522 & .489 \\ -.522 & 3.252 & -3.690 \\ .489 & -3.690 & 4.665 \end{pmatrix}$$

with deviance $= 3.195$ ($df = 5$). The deviances and a plot (not presented here) of proportions of beetles killed against dose levels with the estimated proportions from each model superimposed suggest that the model with the quadratic term is a significantly better fit for these data. Suppose the experimenter is inclined to use the simple logistic fit for future data for ease of interpretation and model simplicity or that the adequacy of the model with the quadratic term is itself in doubt. We proceed by estimating the contamination and then smoothing over the design space as discussed in §2. The resulting design, obtained using the parameter estimates $\hat{\beta}_0^{(1)}$ and $\hat{\beta}_1^{(1)}$ as initial guesses, with total number of observations $n = 481$ over an equally spaced grid of $N = 40$ points in $[0, 1]$ is presented in panel (a) of Figure 14. This would be the design of choice if the experimenter were interested in prediction but contemplated the superiority of the model with quadratic term. However, the experimenter can ensure robustness against a broader set of alternatives by taking the contamination to belong to the class $\mathcal{F}$ while assuming an initial multivariate normal prior on the parameter, with mean vector $\left(\hat{\beta}_0^{(1)}, \hat{\beta}_1^{(1)}\right)^T$ and variance-covariance matrix $\Sigma^{(1)}$, and then using the Bayesian paradigm as in Example 5. The loss function becomes the expected value of (3.10), with expectation taken with respect to the multivariate normal prior. The numerical implementation of expectation is done using a quasi-Monte Carlo sampling approach. The design plot is given in Figure 14(b).

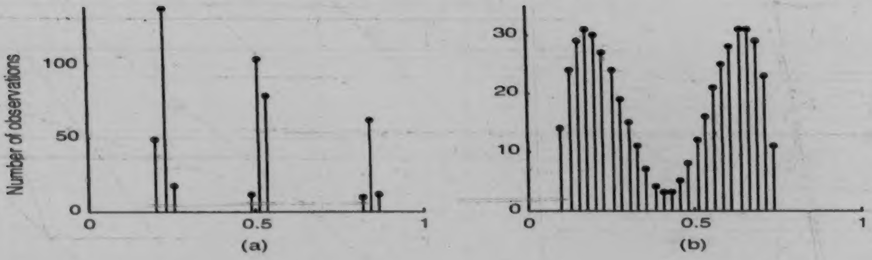

Figure 14: (a) Prediction design when contamination is estimated from initial data. (b) Robust Bayesian prediction design with multivariate normal prior and $\rho = 5$.

3.5 Conclusions

We have investigated integer-valued designs for logistic regression models, using polynomial predictors as specific examples. Our designs are robust against misspecification in the predictor. We have addressed both known and unknown contamination. Previous robustness work done for logistic models has concentrated on the uncertainty of model parameters; in this contribution we have gone further to investigate specific violations in the form of the assumed linear (in the parameters) predictor.

Designs for a specific alternative, for example quadratic versus linear in the independent variable, are quite different from those for broad classes of alternatives. The number of distinct design points is usually not as large in the former case as in the latter. In fact, when the magnitude of the misspecification is minimal the resulting robust design could have about the same number of distinct observation points as its classical counterpart. Nevertheless, the gain in robustness often exceeds the premium paid for robustness - see Table 4.

Designs for a very specific alternative may, however, suffer the same fate as designs assuming the exactness of the fitted model when the alternative itself is not valid. Both take replicates of observations at only a few distinct points, especially when the magnitude of the departure is small. However, when there is a higher degree of certainty in the alternative, these designs could result in substantial gain in robustness. An example of this would be when the experimenter is aware of a more appropriate model but seeks a design that allows for the fitting of a more parsimonious model. Also, designs when there are data to estimate model contamination are quite similar to designs when the exact form of the contamination is known (single alternative). When the information in the initial data is incorporated into the design procedure, as seen in Example 3 above, the robustness of the resulting design could come at a very reduced premium.

In general, we have found there to be increasing numbers of distinct observation sites with increasing model uncertainty. The overall message is consistent with that reported in the model robust design literature for linear models - robust designs can be approximated by placing clusters of observations about the support points for classical designs. However, the nonlinearity of the mean response in logistic design adds a slight twist to the overall message, in that the clusters of observation come with patterns that are determined by the curvature prescribed by the model parameters. More striking is the fact that the all-bias design is non-uniform in logistic regression models - even though the recommended design points are spread over the entire design space, the frequencies of observations vary due to the curvature.

Overall, the design that protects against uncertainty in model parameters (via a Bayesian paradigm) and that which protects against uncertainty in assumed model form could be described as taking observations in clusters. These clusters often come in interesting patterns of curvature prescribed by the nonlinearity of the model - see examples in the previous section. Further work would be required to obtain analytic descriptions of the effect of curvature in this robust approach, or even for the all-bias designs for logistic models. While the focus of the model misspecification reported here is exclusively on linear predictor misspecification, we are currently investigating other forms of misspecification in designing for the broader class of generalized linear models, of which the logistic model is but a special case.

3.6 Appendix: Derivations

Proof of Theorem 3.1: Under conditions as in Fahrmeir (1990) the maximum likelihood estimate $\hat{\boldsymbol{\beta}}$ exists and is consistent, and $\partial l\left(\hat{\boldsymbol{\beta}}\right)/\partial\boldsymbol{\beta}$ is $o_p\left(n^{-1/2}\right)$. The log-likelihood l, the score function and -1 times the second derivative according to the assumed model are

$$l\left(\boldsymbol{\beta}\right)=\sum_{i=1}^{N}\left\{n_i\left[y_i\log\left(\frac{\mu_i}{1-\mu_i}\right)+\log\left(1-\mu_i\right)\right]+\log\binom{n_i}{n_iy_i}\right\},$$

$$\frac{\partial l\left(\boldsymbol{\beta}\right)}{\partial\boldsymbol{\beta}}=\sum_{i=1}^{N}n_i\left(y_i-\mu_i\right)\mathbf{z}\left(\mathbf{x}_i\right),\quad-\frac{\partial^2 l\left(\boldsymbol{\beta}\right)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}=\sum_{i=1}^{N}n_iw_i\mathbf{z}\left(\mathbf{x}_i\right)\mathbf{z}^T\left(\mathbf{x}_i\right).$$

An expansion of $\partial l\left(\boldsymbol{\beta}\right)/\partial\beta_j$ around $\boldsymbol{\beta}_0$ gives

$$\frac{\partial l\left(\boldsymbol{\beta}\right)}{\partial\beta_j}=\frac{\partial l\left(\boldsymbol{\beta}_0\right)}{\partial\beta_j}+\sum_k\left(\beta_k-\beta_{0,k}\right)\frac{\partial^2 l\left(\boldsymbol{\beta}_0\right)}{\partial\beta_j\partial\beta_k}+\frac{1}{2}\sum_k\sum_l\left(\beta_k-\beta_{0,k}\right)\left(\beta_l-\beta_{0,l}\right)\frac{\partial^3 l\left(\boldsymbol{\beta}_*\right)}{\partial\beta_j\partial\beta_k\partial\beta_l},$$

where β_j and $\beta_{0,j}$ are the j^{th} terms of the vectors $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$ respectively and $\boldsymbol{\beta}_*$ is a point on the line segment connecting $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$. If we replace $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$ in this expansion, we obtain

$$\sqrt{n}\sum_k \left(\hat{\beta}_k - \beta_{0,k}\right)\left[\frac{1}{n}\frac{\partial^2 l\left(\boldsymbol{\beta}_0\right)}{\partial\beta_j\partial\beta_k} + \frac{1}{2n}\sum_l \left(\hat{\beta}_l - \beta_{0,l}\right)\frac{\partial^3 l\left(\boldsymbol{\beta}_*\right)}{\partial\beta_j\partial\beta_k\partial\beta_l}\right] = -\frac{1}{\sqrt{n}}\frac{\partial l\left(\boldsymbol{\beta}_0\right)}{\partial\beta_j}.$$

For the logistic likelihood the $\frac{\partial^3 l(\boldsymbol{\beta}_*)}{\partial\beta_j\partial\beta_k\partial\beta_l}$ are bounded, and so, using the consistency of $\hat{\boldsymbol{\beta}}$, we have that

$$\left[\frac{1}{n}\frac{\partial^2 l\left(\boldsymbol{\beta}_0\right)}{\partial\beta_j\partial\beta_k} + \frac{1}{2n}\sum \left(\hat{\beta}_l - \beta_{0,l}\right)\frac{\partial^3 l\left(\boldsymbol{\beta}_*\right)}{\partial\beta_j\partial\beta_k\partial\beta_l}\right] \xrightarrow{p} -H_{jk},$$

where H_{jk} is the $(j,k)^{th}$ element of the matrix $\mathbf{H}_n = -\frac{1}{n}\frac{\partial^2 l(\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} = \mathbf{Z}^T\mathbf{PWZ}$. Thus the limit distribution of $\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$ is that of the solution of the equations

$$\sum H_{jk}\sqrt{n}\left(\hat{\beta}_k - \beta_{0,k}\right) = \frac{1}{\sqrt{n}}\frac{\partial l\left(\boldsymbol{\beta}_0\right)}{\partial\beta_j},$$

i.e. is the limit distribution of $\mathbf{H}_n^{-1}\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}}$. Using the central limit theorem for independent not identically distributed random variables we have that $\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}}$ has a multivariate normal limit distribution with asymptotic mean

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{N} n_i E\left[y_i - \mu_i\left(\boldsymbol{\beta}_0\right)\right]\mathbf{z}\left(\mathbf{x}_i\right) = \sqrt{n}\mathbf{b}$$

and asymptotic covariance matrix $\tilde{\mathbf{H}}_n = \mathbf{Z}^T\mathbf{PW}_T\mathbf{Z}$. From this it follows that $\sqrt{n}\left(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$ is $AN\left(\sqrt{n}\mathbf{H}_n^{-1}\mathbf{b}, \mathbf{H}_n^{-1}\tilde{\mathbf{H}}_n\mathbf{H}_n^{-1}\right)$, as required. □

Proof of Corollary 3.1: First write

$$I = \frac{1}{N}\sum_{i=1}^{N} var\left[\sqrt{n}\mu\left(\hat{\eta}_i\right)\right] + \frac{1}{N}\sum_{i=1}^{N}\left\{E\left[\sqrt{n}\mu\left(\hat{\eta}_i\right)\right] - \sqrt{n}\mu\left(\eta_i + f\left(\mathbf{x}_i\right)\right)\right\}^2.$$

By the δ-method, the first sum is, up to terms which are $o(1)$,

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{d\mu_i}{d\eta_i}\right)^2 var\left[\sqrt{n}\hat{\eta}_i\right] &= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{d\mu}{d\eta_i}\right)^2 \mathbf{z}^T\left(\mathbf{x}_i\right)\mathbf{H}_n^{-1}\tilde{\mathbf{H}}_n\mathbf{H}_n^{-1}\mathbf{z}\left(\mathbf{x}_i\right)\\ &= \frac{1}{N}tr\left[\mathbf{WZH}_n^{-1}\tilde{\mathbf{H}}_n\mathbf{H}_n^{-1}\mathbf{Z}^T\mathbf{W}\right].\end{aligned}$$

Also, on expanding $\mu(\hat{\eta}_i)$ and $\mu(\eta_i + f(\mathbf{x}_i))$ around η_i, we have

$$E\left[\sqrt{n}\mu(\hat{\eta}_i)\right] = \sqrt{n}\mu(\eta_i) + E\left[\sqrt{n}\frac{d\mu}{d\eta_i}(\hat{\eta}_i - \eta_i) + o(\sqrt{n}(\hat{\eta}_i - \eta_i))\right],$$

and

$$\sqrt{n}\mu(\eta_i + f(\mathbf{x}_i)) = \sqrt{n}\mu(\eta_i) + \sqrt{n}\frac{d\mu}{d\eta_i}f(\mathbf{x}_i) + o(\sqrt{n}f(\mathbf{x}_i)).$$

Using an argument similar to that in the proof of Theorem 3.1, we have

$$E\left[\sqrt{n}\mu(\hat{\eta}_i)\right] = \sqrt{n}\mu(\eta_i) + \sqrt{n}\frac{d\mu}{d\eta_i}E(\hat{\eta}_i - \eta_i) + o(1).$$

Thus, the second sum in the expression of I is, up to terms which are $o(1)$,

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\left\{E\left[\sqrt{n}\mu(\hat{\eta}_i)\right] - \sqrt{n}\mu(\eta_i + f(\mathbf{x}_i))\right\}^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{d\mu}{d\eta_i}\right)^2\left\{n \cdot bias^T\left(\hat{\boldsymbol{\beta}}\right)\mathbf{z}(\mathbf{x}_i)\mathbf{z}^T(\mathbf{x}_i)\,bias\left(\hat{\boldsymbol{\beta}}\right) + nf^2(\mathbf{x}_i)\right\} \\
&= \frac{1}{N}\left\{n \cdot \mathbf{b}^T\mathbf{H}_n^{-1}\mathbf{Z}^T\mathbf{W}^2\mathbf{Z}\mathbf{H}_n^{-1}\mathbf{b} - 2n\mathbf{f}^T\mathbf{W}^2\mathbf{Z}\mathbf{H}_n^{-1}\mathbf{b} + n \cdot \mathbf{f}^T\mathbf{W}^2\mathbf{f}\right\},
\end{aligned}$$

reducing to $\frac{n}{N}\left\|\mathbf{W}\left(\mathbf{Z}\mathbf{H}_n^{-1}\mathbf{b} - \mathbf{f}\right)\right\|^2$. □

Proof of Theorem 3.2: Here and elsewhere, in the averaging we will use the identity $\int \mathbf{t}^T\mathbf{t}p(\mathbf{t})\,d\mathbf{t} = (N-p)/(N-p+2)$, which implies that

$$\int \mathbf{t}\mathbf{t}^T p(\mathbf{t})\,d\mathbf{t} = \frac{1}{N-p+2}\mathbf{I}_{N-p}.$$

First use (3.8) and (3.9) to write (3.7) explicitly in terms of $\mathbf{t}$:

$$\mathcal{L}_I(\mathbf{P},\mathbf{f}) = \frac{1}{N}\left\{\begin{array}{l} tr\left[\left(\mathbf{U}^T\mathbf{P}\mathbf{W}\mathbf{U}\right)^{-1}\left(\mathbf{U}^T\mathbf{P}\mathbf{W}_T(\mathbf{t})\,\mathbf{U}\right)\left(\mathbf{U}^T\mathbf{P}\mathbf{W}\mathbf{U}\right)^{-1}\mathbf{U}^T\mathbf{W}^2\mathbf{U}\right] \\ +n\left\|\mathbf{W}\left(\mathbf{U}\left(\mathbf{U}^T\mathbf{P}\mathbf{W}\mathbf{U}\right)^{-1}\mathbf{U}^T\mathbf{P}\left(\boldsymbol{\gamma}_T(\mathbf{t}) - \boldsymbol{\gamma}\right) - \tau\sqrt{N}\tilde{\mathbf{U}}\mathbf{t}\right)\right\|^2 \end{array}\right\}. \tag{A.1}$$

Using (3.9) again we have $\mathbf{W}_T(\mathbf{t}) = \mathbf{W} + \dot{\mathbf{W}}\tau\sqrt{N}\tilde{\mathbf{U}}\mathbf{t} + O(\tau^2)$, where $\dot{\mathbf{W}} = diag\left(w'(\eta_1), \cdots w'(\eta_N)\right)$. Since $\tau^2 = O(n^{-1})$ we obtain $\int \mathbf{W}_T(\mathbf{t})\, p(\mathbf{t})\, d\mathbf{t} = \mathbf{W} + O(n^{-1})$, and so

$$\begin{aligned}&\int tr\left[\left(\mathbf{U}^T\mathbf{PWU}\right)^{-1}\left(\mathbf{U}^T\mathbf{PW}_T(\mathbf{t})\,\mathbf{U}\right)\left(\mathbf{U}^T\mathbf{PWU}\right)^{-1}\mathbf{U}^T\mathbf{W}^2\mathbf{U}\right] p(\mathbf{t})\, d\mathbf{t}\\ &= tr\left[\left(\mathbf{U}^T\mathbf{PWU}\right)^{-1}\left(\mathbf{U}^T\mathbf{W}^2\mathbf{U}\right)\right].\end{aligned}$$

Similarly, we have $\boldsymbol{\gamma}_T(\mathbf{t}) - \boldsymbol{\gamma} = \tau\sqrt{N}\mathbf{W}\tilde{\mathbf{U}}\mathbf{t} + O(\tau^2)$, and so

$$\begin{aligned}&n\left\|\mathbf{W}\left(\mathbf{U}\left(\mathbf{U}^T\mathbf{PWU}\right)^{-1}\mathbf{U}^T\mathbf{P}\left(\boldsymbol{\gamma}_T(\mathbf{t}) - \boldsymbol{\gamma}\right) - \tau\sqrt{N}\tilde{\mathbf{U}}\mathbf{t}\right)\right\|^2\\ &= n\tau^2 N\left\|\mathbf{W}\left(\mathbf{R} - \mathbf{I}\right)\tilde{\mathbf{U}}\mathbf{t}\right\|^2 + O\left(n^{-1/2}\right),\end{aligned}$$

with

$$\begin{aligned}&\int n\left\|\mathbf{W}\left(\mathbf{U}\left(\mathbf{U}^T\mathbf{PWU}\right)^{-1}\mathbf{U}^T\mathbf{P}\left(\boldsymbol{\gamma}_T(\mathbf{t}) - \boldsymbol{\gamma}\right) - \tau\sqrt{N}\tilde{\mathbf{U}}\mathbf{t}\right)\right\|^2 p(\mathbf{t})\, d\mathbf{t}\\ &= \frac{n\tau^2 N \cdot tr\left[\mathbf{W}\left(\mathbf{R} - \mathbf{I}\right)\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T\left(\mathbf{R} - \mathbf{I}\right)^T\mathbf{W}\right]}{N - p + 2}.\end{aligned}$$

The result follows upon noting that $\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T = \mathbf{I} - \mathbf{U}\mathbf{U}^T$ and $\left(\mathbf{R} - \mathbf{I}\right)\mathbf{U} = \mathbf{0}$, and then substituting these integrals into (A.1) and simplifying. ☐

Acknowledgements

The research of both authors is supported by the Natural Sciences and Engineering Research Council of Canada.

References

Abdelbasit, K.M. and Plackett, R.L. (1983), "Experimental Design for Binary Data," *Journal of the American Statistical Association*, 78, 90-98.

Adewale, A. and Wiens, D.P. (2006), "New Criteria for Robust Integer-Valued Designs in Linear Models," *Computational Statistics and Data Analysis*, in press.

Atkinson, A.C. and Haines, L.M. (1996), "Designs for Nonlinear and Generalized Linear Models," in: *Handbook of Statistics*, ed. S. Ghosh and C.R. Rao, (13) pp. 437 - 475.

Bliss, C.I. (1935), "The Calculation of the Dose-mortality Curve. *Annals of Applied Biology*, 22, 134 - 167.

Box, G.E.P., and Draper, N.R. (1959), "A basis for the selection of a response surface design," *Journal of the American Statistical Association*, 54, 622 - 654.

Burridge, J. and Sebastiani, P. (1994), "D-optimal Designs for Generalized Linear Models with Variance Proportional to the Square of the Mean," *Biometrika*, 81, 295 - 304.

Chaloner, K. and Larntz, K. (1989), "Optimal Bayesian Design Applied to Logistic Regression Experiments," *Journal of Statistical Inference and Planning*, 21, 191 - 208.

Chaudhuri, P. and Mykland, P. (1993), "Nonlinear experiments: Optimal Design and Inference Based on Likelihood," *Journal of the American Statistical Association*, 88, 538-546.

Chernoff, H. (1953), "Locally Optimal Designs for Estimating Parameters," *Annals of Mathematical Statistics*, 24, 586-602.

Dette, H. and Wong, W.K. (1996), "Optimal Bayesian Designs for Models with Partially Specified Heteroscedastic structure," *Annals of Statistics*, 24(5), 2108 - 2127.

Dette, H., Haines, L. and Imhof, L. (2003), "Bayesian and maximin optimal designs for heteroscedastic regression models," *The Canadian Journal of Statistics*, 33(2), 221-241.

Fahrmeir, L. (1990), "Maximum Likelihood Estimation in Misspecified Generalized Linear Models," *Statistics*, 21 (4), 487-502.

Fang, K.-T. and Wang, Y. (1994), *Number-Theoretic Methods in Statistics*, Chapman & Hall, London.

Fang, Z., and Wiens, D.P. (2000), "Integer-valued, Minimax Robust Designs for Estimation and Extrapolation in Heteroscedastic, Approximately Linear Models," *Journal of the American Statistical Association*, 95, 807-818.

Fedorov, V.V. (1972), *Theory of Optimal Experiments*, Academic Press, New York.

Ford, I. and Silvey, S.D. (1980), "A sequentially constructed design for estimating a nonlinear parametric function," *Biometrika*, 67, 381-388.

Ford, I, Titterington, D.M. and Kitsos, C.P. (1989), "Recent Advances in Nonlinear Experimental Design," *Technometrics*, 31, 49-60.

Ford, I. Torsney B. and Wu, C.F.J. (1992), "The Use of a Canonical Form in the Construction of Locally Optimal Designs for Nonlinear Problems," *Journal of* *the Royal Statistical Society* Series B, 54, 569-583.

Kitsos, C.P., Titterington, D.M. and Torsney, B (1988), "An Optimal Design Problem in Rhythmometry," *Biometrics*, 44, 657-671.

Läuter, E. (1974), "Experimental Design in a Class of Models," *Mathematische Operationsforschung und Statistik* 5, 379-396.

Läuter, E. (1976), "Optimal Multipurpose Designs for Regression Models," *Mathematische Operationsforschung und Statistik* 7, 51-68.

Li, K.C., and Notz, W. (1982), "Robust Designs for Nearly Linear Regression," *Journal of Statistical Planning and Inference*, 6, 135-151.

Marcus, M.B., and Sacks, J. (1976), "Robust Designs for Regression Problems," in: *Statistical Theory and Related Topics II*, ed. S.S. Gupta and D.S. Moore, New York: Academic Press, pp. 245-268.

McCullagh, P. and Nelder, J.A. (1983), *Generalized Linear Models*, Chapman & Hall/CRC.

Meyer, R.K., and Nachtsheim, C.J. (1988), "Constructing Exact D-Optimal Experimental Designs by Simulated Annealing," *American Journal of Mathematical and Management Sciences*, 3 & 4, 329-359.

Minkin, S. (1987), "Optimal Design for Binary Data," *Journal of the American Statistical Association*, 82, 1098-1103.

Montgomery, D.C., Peck, E.A., and Vining, G.G (2001), *Introduction to Linear Regression Analysis*, John Wiley, NY.

Pregibon, D. (1980), "Goodness of Link Selection in Generalized Linear Models," *Applied Statistics*, 29, 15 - 24.

Sinha, S. and Wiens, D.P. (2002), "Robust Sequential Designs for Nonlinear Regression," *The Canadian Journal of Statistics*, 30, 601-618.

Sitter, R.R. (1992), "Robust Designs for Binary Data," *Biometrics*, 48, 1145-1155.

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-25.

Wiens, D.P. (1992), "Minimax Designs for Approximately Linear Regression," *Journal of Statistical Planning and Inference*, 31, 353-371.

Wiens, D.P. and Zhou, J. (1999), "Minimax Designs for Approximately Linear Models with AR(1) Errors," *The Canadian Journal of Statistics*, 27 (4), 781-794.

CHAPTER IV

ROBUST DESIGNS FOR BINOMIAL MODELS WITH OVERDISPERSION AND MISSPECIFIED LINK FUNCTIONS

Abstract We have developed criteria that generate robust designs which insure against possible misspecifications in logistic models. We construct designs that protect against overdispersion and misspecified link functions. The problem of overdispersion is addressed by incorporating the variance function prescribed by a superior model similar to a logistic mixed model. The logistic model corresponds to the canonical link for the binomial distribution. Our robust approach to design deals with the problem of link misspecification by adopting a parameterized generalized family of link functions encompassing the logistic link and other alternatives. The design criterion is the average mean squared error of predictions over a finite design space which depends on unknown model coefficients, overdispersion and link parameters. We use the "minave" approach, which requires the specification of ranges for unknown parameters, to construct robust designs. Examples of minave optimal designs are presented, including an example on the toxicity of ethylene oxide to grain beetles.

4.1 Introduction

In this article we present robust designs for binomial models. The literature is replete with works on model-robust designs for linear models but there is little work done

---

[3] Submitted for publication.

on model-robust designs in the context of binomial models. A possible reason for the dearth of work in this area is explained by the complexity of the design problem even when the assumed model is exactly correct. Contributions to the experimental design literature with the notion that the assumed logistic model is exactly correct include those of Abdelbasit and Plackett (1983), Minkin (1987), Ford, Torsney and Wu (1992), Chaudhuri and Mykland (1993), Burridge and Sebastiani (1994), and Atkinson and Haines (1996) but to name a few. The general approach to design is to optimize criteria which are real-valued functions of the information matrix of the model parameters. One main difficulty is the dependency of design criteria on unknown model parameters. A traditional approach around this difficulty is to use best guesses for parameter values. This was termed locally optimal designs by Chernoff (1953). Authors such as Chaloner and Larntz (1989) and Dette and Wong (1996), have used a Bayesian paradigm - assuming a prior distribution on the unknown parameters. There have also been minimax (or maximin) proposals for robustification of the uncertainties in model parameters (see Sitter, 1992, King and Wong, 2000 and Biedermann, Dette and Pepelyshev, 2004).

In an article by Ford, Titterington and Kitsos (1989) on nonlinear designs, it was stated that when the assumed model is seriously in doubt, designs based on such model may be grossly inappropriate. Ford et al. are probably just alluding to what is known in the context of linear models through the pioneering work of Box and Draper (1959) and subsequent work by Huber (1975), Marcus and Sacks (1976), Li and Notz (1982) and Wiens (1990,1992, 1998).

There are three ways a binomial model can be potentially misspecified when viewed as a generalized linear model. First, the covariates included in the systematic component of the model, the linear predictor, may not reflect the influence of covariates correctly. This may be due to the use of a wrongly specified functional form of the covariates in the model or an omission of essential covariates. Second, the link function might not be accurate. For example, the use of the logit link which is the canonical link for the binomial distribution when in fact the complementary log-log link or the probit link is more appropriate. A third source of misspecification in binomial model is extra-binomial variation. Extra-binomial variation is a situation whereby the nominal variance prescribed by the binomial distribution does not correctly account for observed variability in the data. Overdispersion is the most common form of extra-binomial variation, it occurs when the data exhibit variability that exceeds that prescribed by the binomial distribution. Underdispersion is the opposite, but it is not as common as overdispersion (see Dean, 1992).

Adewale and Wiens (2005) considered designs for logistic models with an eye on possible misspecification in the fixed effects specified through the linear predictor. Their work addressed the problem of designing for logistic models when the linear predictor is possibly misspecified. This article extends the work of Adewale and Wiens (2005) to designing for logistic models when there is overdispersion and when the logit link is possibly inadequate. Our approach to robust design is to regard the true model as belonging to a broad class of models accommodating overdispersion and several link functions other than the logit link. We compute the mean squared error of predictions as a composition of variance error as well as bias due to model misspecification. Our robust designs optimize certain functions of this mean squared

Table 8: Toxicity of Ethylene Oxide to the Calandra Granaria.

Dose x_i ($\log_{10} C_2H_4O\ mg/100\ ml$)	# of observations n_i	# dead n_iy_i
.394	30	23
.391	30	30
.362	31	29
.322	30	22
.314	26	23
.260	27	7
.225	31	12
.199	30	17
.167	31	10
.033	24	0

Table 9: Bliss's Beetle mortality data.

Dose, x_i (mgl^{-1})	49.06	52.99	56.91	60.84	64.76	68.69	72.61	76.54
Number of beetles, n_i	59	60	62	56	63	59	62	60
Number killed, n_iy_i	6	13	18	28	52	53	61	60

error of predictions.

Example 1.1 (Overdispersion illustrated): Busvine (1938) presented a data set for the mortality of grain beetle (*Calandra Granaria)* after exposure to ethylene oxide (C_2H_4O). A total of 290 grain beetles were exposed to 10 different levels of concentrations of C_2H_4O and the proportion killed at each concentration after a 1-hour period was recorded (Table 8). When the logistic linear model was fitted to the data we obtained the parameter estimates $\hat{\beta}_0 = -3.443$ and $\hat{\beta}_1 = 14.440$ and the deviance statistic is 33.24 ($8\,d.f.$). The deviance statistic suggests lack of fit. The binomial model with other link functions offers no improved fit. A plot of proportion of mortality versus dosage with the fitted logistic model superimposed suggests that the lack of fit is possibly due to extra variation around the fitted mean response line (see Figure 15).

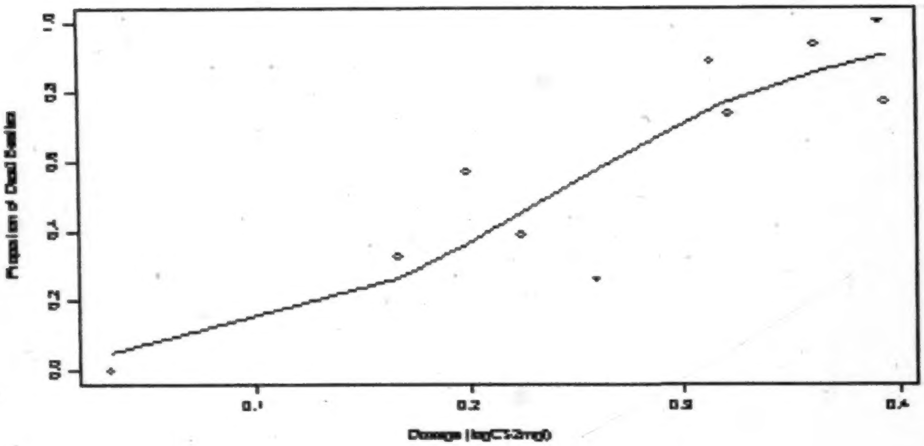

Figure 15: Mortality of Calandra Granaria on exposure to ethylene oxide

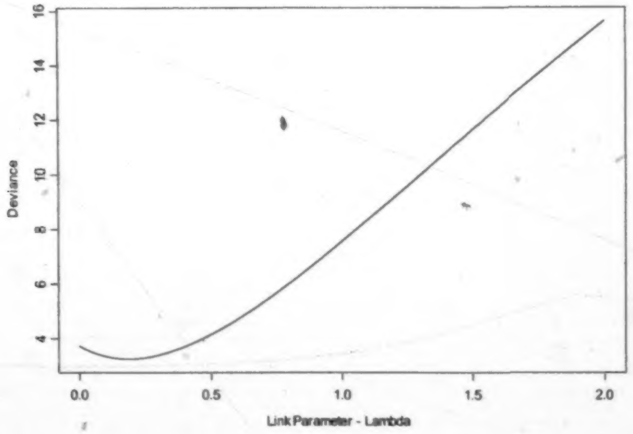

Figure 16: Deviance versus λ for the beetle mortality data

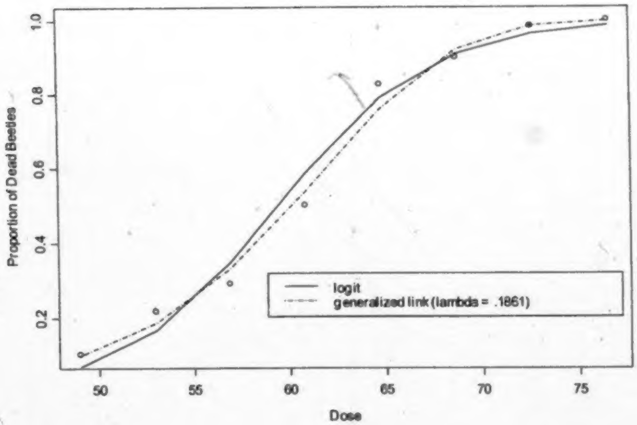

Figure 17: Proportion of dead beetles versus dose with logit link and $\lambda = .1861$-link superimposed

Example 1.2 (Link misspecification illustrated): Table 9 presents Bliss's (1935) data on the numbers of dead beetles after five hours exposure to gaseous carbon disulphide at various concentrations. We fit the binomial model corresponding to a range of values of λ for a generalized link function (defined by 4.5 in §4.2.2). In particular, we fit the model for a grid of λ values in $[0, 2]$. A plot of deviance versus λ values is presented in Figure 16. Using the deviance as the model fit criterion, $\lambda = 0.1861$ corresponds to the model with the minimum deviance. Figure 17 shows that the model with the link parameter $\lambda = 0.1861$ provides an improved fit to the data than the logistic model with $\lambda = 1$. The parameter estimates for the logistic model are $\hat{\beta}_0 = -14.808$ and $\hat{\beta}_1 = 0.249$ and the corresponding estimates for the model with $\lambda = 0.1861$ are $\hat{\beta}_0 = -10.782$ and $\hat{\beta}_1 = 0.174$. In an investigation of designs for nonlinear models, Sinha and Wiens (2002) asserted that "Although the theoretical response functions are very similar in shape, and possibly indistinguishable if noisy data must be relied upon, the appropriate designs can be quite dissimilar." Similarly, since the link function determines the response function in a generalized linear model, the misspecification of the link function could be consequential on the designs.

4.2 Statistical Models

We consider the case of an experimenter with a finite set $\mathcal{S} = \{\mathbf{x}_i\}_{i=1}^{N}$ of possible design points whose interest is to choose n, not necessarily distinct, points at which to observe binomial proportions, say Y. The experimenter makes $n_i \geq 0$ observations at x_i such that $\sum_{i=1}^{N} n_i = n$. The design problem is to choose $n_1, ..., n_N$ in an optimal manner and with an eye on possible model misspecification. The objective then is to choose a probability distribution $\{p_i\}_{i=1}^{N}$, with $p_i = n_i/n$, on the design space $\mathcal{S} = \{\mathbf{x}_i\}_{i=1}^{N}$, in some robust and optimal manner.

The experimenter intends to fit the logistic model with the mean response

$$\mu = \mu(\eta) = \frac{\exp(\eta)}{1 + \exp(\eta)}, \tag{4.1}$$

for the linear predictor $\eta = \mathbf{z}^T(\mathbf{x})\boldsymbol{\beta}$ with $\mathbf{z}(\mathbf{x})$ as a vector of predictors. However, the experimenter seeks protection against possible misspecifications in the assumed model form. As highlighted in the introduction, the following are the kinds of departure from the assumed model that the experimenter is worried about:

1. The true linear predictor is $\eta = \mathbf{z}^T(\mathbf{x})\boldsymbol{\beta} + f(\mathbf{x})$ where the function $f(\mathbf{x})$ is some contamination function accounting for additional effects of covariates.

2. The binomial variance, $var\left(Y_i|\mathbf{x}_i\right) = \frac{\mu(\eta_i)(1-\mu(\eta_i))}{n_i}$, assumed by the logistic model (4.1) do not correctly reflect the variability in the anticipated data.

3. The true link function is different from the logit link function, $\eta_i = \log\left(\frac{\mu_i}{1-\mu_i}\right)$, assumed by the logistic model (4.1).

The subject of designing for logistic models with possible misspecification in the linear predictor has been treated in Adewale and Wiens (2005). Adewale and Wiens assumed that the contamination function $f\left(\mathbf{x}\right)$ belongs to a contamination neighbourhood $\mathcal{F}$,

$$\mathcal{F} = \left\{ \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}\left(\mathbf{x}_i\right) f\left(\mathbf{x}_i\right) = 0,\ \frac{1}{N}\sum_{i=1}^{N} f^2\left(\mathbf{x}_i\right) \leq \tau^2, \text{ with } \tau^2 = O\left(n^{-1}\right) \right\}.$$

They noted that the first condition in $\mathcal{F}$ ensures identifiability of the linear predictor and the second condition ensures that the bias engendered by the misspecified linear predictor remains bounded. The average (over the design space $\mathcal{S}$) mean squared error in predicting the mean response, $\mu\left(\eta\right)$ was adopted as loss function. This loss function depends on the unknown contamination function $f\left(\mathbf{x}\right)$. In order to eliminate this dependency of the loss function on the unknown contamination function $f\left(\mathbf{x}\right)$, they used the average of the loss function over the contamination neighbourhood $\mathcal{F}$ as the design criterion. The focus of this article is to treat the problem of designing for logistic models when the concern is about possible misspecification of the second and third kinds.

4.2.1 Overdispersion Model

Overdispersion is a very common phenomenon in data involving proportions but most of the attention it has received in the literature has been in the context of data analysis. Modelling overdispersion has been considered by, among others, Pierce and Sands (1975), Crowder (1978) and Williams (1982). However, this subject has received very scant attention in the regression design literature. Minkin (1993) considers overdispersion when constructing designs for Poisson data. In order to construct robust designs that give protection for overdispersion we assume a superior model that accommodates overdispersion as the true model. In addition, the assumed true model is also required to have the fitted logistic model as a special case. Here we adopt a model like the logistic-mixed model which accommodates overdispersion by incorporating an additive random component to the linear predictor. Thus the problem of designing for logistic models with overdispersion is somewhat a problem of designing for logistic models with misspecified linear predictors. The true linear predictor has two parts - the fixed part representing the effects of covariates and a random part capturing overdispersion. Formally, the true but unknown model belongs to a class of alternative models defined as follows:

$$\begin{aligned} \mu_i &= \mu(\eta_i) = \frac{\exp(\eta_i)}{1+\exp(\eta_i)}, \\ \eta_i &= \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta}_0 + \phi v_i, \text{ with} \\ E(v_i) &= 0 \text{ and } var(v_i) = 1. \end{aligned} \tag{4.2}$$

Thus, the true model has $Y_i|v_i \sim \frac{1}{n_i} binomial(n_i, \mu(\eta_i))$, $\eta_i = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta}_0 + \phi v_i$, with $E(v_i) = 0$ and $var(v_i) = 1$, so that, $E(Y_i|v_i) = \mu(\eta_i)$, $var(Y_i|v_i) = \frac{E(Y_i|v_i)(1-E(Y_i|v_i))}{n_i}$.

We take the following Taylor's expansion:

$$\begin{aligned} E(Y_i|v_i) &= \frac{\exp(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta})}{1+\exp(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta})} + \mu'(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta})\phi v_i + \mu''(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta})\phi^2 v_i^2 + ... \\ &= \mu(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta}) + \mu'(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta})\phi v_i + O(\phi^2). \end{aligned}$$

Let $w(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta}) = \mu'(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta})$ (which is, $= \mu(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta})(1-\mu(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta}))$)

$$var(Y_i|v_i) = \frac{1}{n_i}\left\{ \begin{array}{c} w(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta}) + (1-\mu(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta}))\, w(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta})\phi v_i \\ -\mu(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta})\, w(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta}) - w^2(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta})\phi^2 v_i^2 + O(\phi^2) \end{array} \right\}.$$

Therefore,

$$E\{var(Y_i|v_i)\} = \frac{1}{n_i}\left\{w(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta})\left[1 - w(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta})\phi^2\right]\right\} + O(\phi^2), \text{ and}$$

$$var\{E(Y_i|v_i)\} = w^2(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta})\phi^2 + O(\phi^3).$$

Using the identities

$$E(Y_i) = E\{E(Y_i|v_i)\},$$

and

$$var(Y_i) = var\{E(Y_i|v_i)\} + E\{var(Y_i|v_i)\},$$

the true mean response and true variance are given by

$$\begin{aligned} E(Y_i) &= \mu(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta}) + O(\phi^2), & (4.3) \\ var(Y_i) &= \frac{1}{n_i} w(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta})\left\{1 + (n_i - 1)\, w(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta})\phi^2\right\} + O(\phi^2) & (4.4) \end{aligned}$$

4.2.2 True Link Model

Ponce de Leon and Atkinson (1993) investigated designing optimal experiments for the choice of link function for a binary data model. They assumed a framework where interest is in the estimation of the link function as well as the model coefficients when the link function belongs to the same parameterized generalized link family of interest in this article. Thus their approach fits into the context of classical approach to regression design in which the experimenter takes the assumed model to be exact. The difference is that the link function is rather added as an extra-parameter to be estimated from the data. Biedermann, Dette and Pepelyshev (2004) introduced a robust approach in which the experimenter is considering a finite set of plausible link functions with the uncertainties in the suitability of each of them quantified with known probabilities. The probabilities reflect the preferences the experimenter attaches to each link function. The resulting robust criterion is a weighted average of the respective criterion corresponding to each link function. We take a different approach here. Given that the experimenter intends to fit the logistic model, binomial model with the logit link, we propose designs that protect against the possibility that the logit link is not exactly correct.

The framework for the robust design in this work is that the true but unknown link function as well as the fitted logit link belong to the generalized family of link functions

$$g(\mu, \lambda) = \log\left[\left\{\left(\frac{1}{1-\mu}\right)^{\lambda} - 1\right\}/\lambda\right], \quad (\lambda \geq 0), \tag{4.5}$$

parameterized by λ. The true link function corresponds to an unknown value of the link parameter that might be different from that which corresponds to the logistic model. This generalized family of links encompasses the logit and the complementary log-log links as special cases. The logit link corresponds to $\lambda = 1$ while the complementary log-log corresponds to $\lim_{\lambda \to 0}$. We proceed by casting the problem rather as a linear predictor misspecification problem. Adewale and Wiens (2005) have investigated robust designs for logistic model with possibly misspecified linear predictor.

In generalized linear modelling the link function connects the systematic component (the linear predictor) of the model to the mean response via

$$\eta = g(\mu, \lambda),$$

where η is the linear predictor representing the effects of covariates in the model on a linear scale. The link function determines the distribution of the mean response, for example if $g(\mu, \lambda)$ is the logit function, the mean response takes the logistic distribution function and if $g(\mu, \lambda)$ is the complementary log-log link it takes the extreme value distribution. We considered the following Taylor's expansion of (4.5) about the parameter value, $\lambda = 1$ corresponding to the logit link that the experimenter contemplates fitting. That is,

$$\eta = \log\left(\frac{\mu}{1-\mu}\right) + h(\mathbf{x}; \lambda) \tag{4.6}$$

with μ given by (4.1) and

$$h(\mathbf{x}; \lambda) = \left.\frac{\partial g}{\partial \lambda}\right|_{\lambda=1} (\lambda - 1) + o(\lambda - 1). \tag{4.7}$$

Thus the link misspecification problem is again cast as a linear predictor misspecification problem. The true mean response, $\mu_{T,i}$ is given by (4.1) with the linear predictor given by

$$\eta_i = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta}_0 + f(\mathbf{x}_i, \lambda), \tag{4.8}$$

where the contamination function is given by

$$f(\mathbf{x},\lambda) = -h(\mathbf{x},\lambda). \tag{4.9}$$

Pregibon (1980) in his work on goodness of link tests for generalized linear models introduced this concept of casting a misspecified link function as an additive contamination in the linear predictor. To this approach he commented that "The fact that the difference in link functions appears on the right side of the link defining equation should not be disturbing-indeed, this corresponds to the fact that the wrong link function is a systematic mis-specification of the model."

4.3 Loss Function and Algorithm

4.3.1 Design Criterion

We use the normalized average mean squared error (AMSE) I of the response prediction $\mu(\hat{\eta}_i)$, with $\hat{\eta}_i = \mathbf{z}^T(\mathbf{x}_i)\hat{\boldsymbol{\beta}}$, as the design criterion,

$$I \triangleq \frac{n}{N}\sum_{i=1}^{N} E\left[\left\{\mu(\hat{\eta}_i) - \mu(\eta_i + f(\mathbf{x}_i))\right\}^2\right]. \tag{4.10}$$

In order to express (4.10), in terms of average variance of predictions and average bias of predictions we employed the following Theorem 1 and Corollary 1 from Adewale and Wiens (2005).

Theorem 4.1 (Adewale and Wiens, 2005) *Define*

$$\begin{aligned} w_i &= \frac{d\mu_i}{d\eta_i} = \mu_i(1-\mu_i) = \frac{1}{4}sech^2\left(\frac{\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta}_0}{2}\right), && (4.11)\\ \mu_{T,i} &= \frac{\exp(\eta_{T,i})}{1+\exp(\eta_{T,i})},\ \eta_{T,i} = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta}_0 + f(\mathbf{x}_i,\lambda), && (4.12)\end{aligned}$$

and let $\mathbf{Z}$ *be the* $N \times p$ *matrix with rows* $\mathbf{z}^T(\mathbf{x}_i)$. *Let* $\boldsymbol{\gamma}$ *and* $\boldsymbol{\gamma}_T$ *be the* $N \times 1$ *vectors with elements* μ_i *and* $\mu_{T,i}$ *respectively. Let* $\mathbf{P}$, $\mathbf{W}$ *and* $\mathbf{W}_T$ *be the* $N \times N$ *diagonal matrices with diagonal elements* n_i/n, w_i *and* $w_{T,i} = \mu_{T,i}\left(1 - \mu_{T,i}\right)$ *respectively. Finally, define* $\mathbf{b} = \mathbf{Z}^T\mathbf{P}\left(\boldsymbol{\gamma}_T - \boldsymbol{\gamma}\right)$, $\mathbf{H}_n = \mathbf{Z}^T\mathbf{PWZ}$, $\tilde{\mathbf{H}}_n = \mathbf{Z}^T\mathbf{PW}_T\mathbf{Z}$, *and* $\mathbf{G}_n = \mathbf{Z}^T\mathbf{W}_T\mathbf{P}\left(n\mathbf{P} - \mathbf{I}\right)\mathbf{W}_T\mathbf{Z}$. *The asymptotic bias and asymptotic covariance matrix of the maximum likelihood estimator* $\hat{\boldsymbol{\beta}}$ *of the model parameter vector* $\boldsymbol{\beta}$ *from the misspecified model are*

$$\begin{aligned} bias\left(\hat{\boldsymbol{\beta}}\right) &= E\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) = \mathbf{H}_n^{-1}\mathbf{b} + o\left(n^{-1/2}\right), & (4.13) \\ cov\left(\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)\right) &= \mathbf{H}_n^{-1}\tilde{\mathbf{H}}_n\mathbf{H}_n^{-1} + \mathbf{H}_n^{-1}\mathbf{G}_n\mathbf{H}_n^{-1} + o(1), & (4.14) \end{aligned}$$

respectively.

Proof: Theorem 1 of Adewale and Wiens (2005) with the expressions for the true mean and true variance modified according to the derivations in §4.2.1 and §4.2.2.

Corollary 4.1 (Adewale and Wiens, 2005) *The AMSE has the asymptotic approximation* $I = \mathcal{L}_I\left(\mathbf{P}; \lambda, \phi, \boldsymbol{\beta}\right) + O\left(\phi^2\right) + o(1)$, *where*

$$\mathcal{L}_I\left(\mathbf{P}; \lambda, \phi, \boldsymbol{\beta}\right) = \frac{1}{N}\left\{ \begin{array}{c} tr\left[\mathbf{WZH}_n^{-1}\tilde{\mathbf{H}}_n\mathbf{H}_n^{-1}\mathbf{Z}^T\mathbf{W}\right] \\ +\phi^2 tr\left[\mathbf{WZH}_n^{-1}\mathbf{G}_n\mathbf{H}_n^{-1}\mathbf{Z}^T\mathbf{W}\right] + n\left\|\mathbf{W}\left(\mathbf{ZH}_n^{-1}\mathbf{b} - \mathbf{f}\right)\right\|^2 \end{array} \right\} \tag{4.15}$$

for $\mathbf{f} = \left(f\left(\mathbf{x}_1; \lambda\right), \ldots, f\left(\mathbf{x}_N; \lambda\right)\right)^T$ *where* $f\left(\mathbf{x}_i; \lambda\right)$ *is given by (4.9).*

Proof: Corollary 1 of Adewale and Wiens (2005).

The loss function (4.15) depends on the unknown model parameter vector $\boldsymbol{\beta}$, the overdispersion parameter ϕ, and the link parameter λ. Our approach to dealing with this parameter dependency requires parameter spaces Θ, Φ, and Λ of plausible values. Letting $\Omega = [\Theta \times \Phi \times \Lambda]$, the design criterion is then the average of the loss function over the given joint parameter space Ω. The averaging was carried out over a uniformly scattered set of points obtained on Ω using number-theoretic method or quasi-Monte Carlo method (Fang and Wang, 1994). Given the parameter space Ω, we seek designs that minimize the average (over the space Ω) of the loss function:

$$\mathcal{L}_{ave}(\mathbf{P}) = aver_{\Omega}\mathcal{L}_I(\mathbf{P}; \lambda, \phi, \boldsymbol{\beta}), \tag{4.16}$$

through the matrix $\mathbf{P} = diag(n_1/n, \cdots, n_N/n)$. The minimization of (4.16) with respect to $\mathbf{P}$ is undertaken using a modification of the simulated annealing algorithm. It suffices to note here that the averaging in (4.16) is implemented as an integral part of the simulated annealing process - for every new design generated the value of loss $\mathcal{L}_I(\mathbf{P}; \lambda, \phi, \boldsymbol{\beta})$ is calculated for every point $(\lambda, \phi, \boldsymbol{\beta})$ in Ω. The value of $\mathcal{L}_{ave}(\mathbf{P})$ is the average of $\mathcal{L}_I(\mathbf{P}; \lambda, \phi, \boldsymbol{\beta})$ over the points in the parameter space Ω. The resulting designs are termed minave designs.

4.3.2 Algorithm: Modified Simulated Annealing

We consider models with p regressors $(z_1(x), ..., z_p(x))^T$ where $x \in [a, b]$. The design space is the set $\mathcal{S} = \left\{x_i = a + \frac{(b-a)(i-1)}{N-1}\right\}_{i=1}^{N}$ of equally spaced points in $[a, b]$. Given the desired number of observations (n) to be taken and the number of points in the design space (N), we seek designs that minimize the loss function (4.16). Our approach gives the experimenter the flexibility to request a symmetric design about the point $(a + b)/2$ or an asymmetric design.

A modification of the simulated annealing algorithm is employed to search for optimal designs. The simulated annealing algorithm seeks to assign integers $n_i \geq 0$ to each of the design points x_i in such a way that $\mathcal{L}_{ave}$ is a minimum. Simulated annealing is a direct search optimization algorithm which has been quite successful at finding the global extremum of a function, possibly non-smooth, that has many local extrema. The unmodified version of the algorithm consists of the following three steps:

1. A specification of an initial design, say $\mathbf{P}_0$,
2. the second step is the random choice of a new design, say $\mathbf{P}_{new}$, from the optimization space, and lastly,
3. a prescription of the basis of acceptance or rejection of the new design: accept new design with probability π, defined as

   $$\pi = \begin{cases} 1 & \text{if } \Delta\mathcal{L}_{ave} \leq 0, \\ \exp\left(-\Delta\mathcal{L}_{ave}/T\right) & \text{if } \Delta\mathcal{L}_{ave} \geq 0, \end{cases}$$

   where $\Delta\mathcal{L}_{ave} = \mathcal{L}_{ave}\left(\mathbf{P}_{new}\right) - \mathcal{L}_{ave}\left(\mathbf{P}_0\right)$ and T is a tuning parameter. In our implementation we modify step 3 as discussed below.

If $n < N$, then the initial design is chosen to be the design assigning one observation to each of n randomly chosen points. If $n \geq N$, we randomly distribute n observations to all N points. If interest is in symmetric designs we randomly assign one observation to each of $\lfloor \frac{n}{2} \rfloor$ randomly chosen points in $[a, (a+b)/2)$ when $n < N$ and randomly distribute $\lfloor \frac{n}{2} \rfloor$ observations to the $\lfloor \frac{N}{2} \rfloor$ points in $[a, (a+b)/2)$ when

$n \geq N$. If n is odd then N has to be odd for symmetry, in this case we assign the extra observation to the point $\left\{\frac{a+b}{2}\right\}$. A symmetric initial design is obtained by assigning the number of observations for locations in $[a, (a+b)/2)$ to their corresponding mirror image about the point $\left\{\frac{a+b}{2}\right\}$. This completes the first step of the simulated annealing algorithm.

To generate a new design, we use the perturbation scheme presented by Fang and Wiens (2000). Define $\mathbf{v}$ to be the $N \times 1$ current allocation vector. For symmetric designs redefine $\mathbf{v}$ to be the $[N/2] \times 1$ vector consisting of the initial segment $(n_1, ..., n_{[N/2]})$ of the current allocation vector. Define

$$J_+ = \{i | v_i > 0\}, J_0 = \{i | v_i = 0\}$$

with cardinalities $j_+ \geq 1$ and j_0. If $j_+ \geq 2$, generate a Bernoulli random variable with success probability $j_0/(j_0 + j_+)$, choose two indices (t_1, t_2) from J_+, at random without replacement, choose an index t_0 from J_0, at random and modify the selected components of $\mathbf{v}$ as follows:

$$v_{t_0} = v_{t_0} + B,\ v_{t_1} = v_{t_1} - 1 \text{ and } v_{t_2} = v_{t_2} + 1 - B. \tag{4.17}$$

If $j_+ = 1$, choose t_0 from J_0 at random, let t_1 be the index in the singleton set J_+, and then replace (4.17) by

$$v_{t_0} = v_{t_0} + 1,\ v_{t_1} = v_{t_1} - 1.$$

This completes the second step for general designs. For symmetric designs, we complete the modify the scheme as follows. If N is even, then let $\mathbf{n} = (n_1, ..., n_N) = (v_1, ..., v_{N/2}, v_{N/2}, ..., v_1)$. If N is odd, then generate a uniform random variable u.

If $u < 1/N$, with probability 1/2, increase $n_{[N/2]+1}$ by 2 then randomly and symmetrically reduce the remaining n_i by 2; with probability 1/2 reduce $n_{[N/2]+1}$ by 2 then randomly and symmetrically increase the remaining n_i. This step is omitted if $n_{[N/2]+1} < 2$. We then construct $\mathbf{n}$ as described above, with the inclusion of the new frequency $n_{[N/2]+1}$.

The third step is modified as follows. Denote the maximum number of iterations by *maxit* and current iteration by *iter*. Accept the new design with probability π, defined as

$$\pi = \begin{cases} 1 & \text{if } \Delta\mathcal{L}_{ave} \leq 0, \\ uniform\,(.5,.9) & \text{if } \Delta\mathcal{L}_{ave} \geq 0 \text{ and } iter < maxit/16 \\ uniform\,(.25,.5) & \text{if } \Delta\mathcal{L}_{ave} \geq 0 \text{ and } maxit/16 \leq iter < maxit/8 \\ uniform\,(0,.1) & \text{if } \Delta\mathcal{L}_{ave} \geq 0 \text{ and } maxit/8 \leq iter < maxit{*}3/4 \\ 0 & \text{if } \Delta\mathcal{L}_{ave} \geq 0 \text{ and } maxit{*}3/4 \leq iter < maxit \end{cases} .$$

Thus a favourable new design ($\Delta\mathcal{L}_{ave} \leq 0$) is accepted with certainty and an unfavourable new design is accepted according to a separate Bernoulli experiment with success probability chosen such that initially such an unfavourable design is accepted with probability satisfying the inequality $.5 < \pi < .9$. This follows a suggestion from Bohachevsky, Johnson and Stein (1986), that the tuning parameter T in step 3 of the original simulated annealing be chosen such that $.5 < \exp\left(-\Delta\mathcal{L}_{ave}/T\right) < .9$. The probability of acceptance of a detrimental new design is progressively decreased to ensure that the process settles at a global minimum. In an implementation of the original simulated annealing algorithm Fang and Wiens (2000) decrease T by a factor of .9 after each 100 iterations and Adewale and Wiens (2006) decrease T by a factor of .95 after every 20^{th} iteration in some other implementation. In the implementations we considered in section §4.4 and §4.5, we found that various choice of parameter space Ω requires different scheme for decreasing T. We thus adopted this fairly generic scheme which still imitates the characteristics of the original algorithm but obviates the demand to seek a perfect T for different parameter spaces Ω.

4.4 *Examples*

First, we suppose an experimenter has confidence in the adequacy of the logistic model and the specification of the linear predictor for fixed effects but wants protection against overdispersion. The design space is taken to be equally spaced points $\left\{x_i = -1 + \frac{2(i-1)}{N-1}\right\}_{i=1}^{N=41}$ in $[-1, 1]$ and the number of observations to be taken is $n = 200$. We take the range of the model parameters to be $\Theta = [.5, 1.5] \times [2.5, 3.5]$ and construct minave designs for various range of the overdispersion parameter, ϕ. The design for the overdispersion parameter space $[0, .005]$ place 46, 50, 66 and 38 of 200 observations at the points $-.80$, $-.75$, $.10$ and $.15$ respectively. This and other designs corresponding to other ranges of the overdispersion parameter are presented in Figure 18. The range, $\Theta = [.5, 1.5] \times [2.5, 3.5]$ of the model parameters was fixed so as to study the effect of increasing width of the overdispersion parameter space on the resulting designs. The support of the designs broadens over the design space with increasing overdispersion. Similar effects were observed when we considered other model parameter values. In general, the designs protecting against overdispersion are similar to those protecting against systematic misspecification of fixed effects. See Adewale and Wiens (2005). We observed a gradual spread of the design points such that the points that are otherwise the design points for the non-robust design which assumes no misspecification are the foci of the spread of the robust designs.

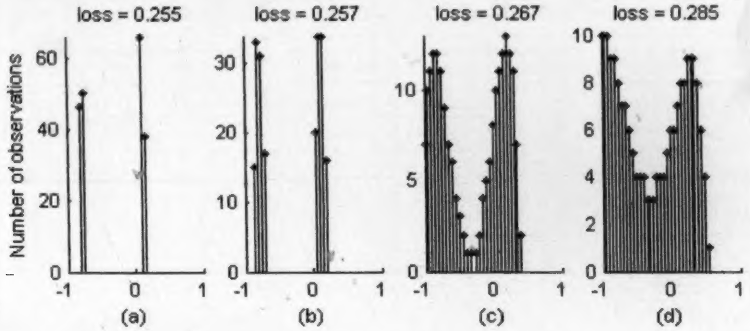

Figure 18: Minave designs when (a) $\Phi = [0, .005]$, (b) $\Phi = [0, .05]$, (c) $\Phi = [0, .25]$, and (d) $\Phi = [0, .5]$ with $\Theta = [.5, 1.5] \times [2.5, 3.5]$ and $\Lambda = \{1\}$.

To evaluate these designs we define efficiency and payoff as,

$$\text{Efficiency} = \left(\frac{aver_{\Theta}\mathcal{L}_I\left(\mathbf{P}_{non-robust}; \lambda = 1, \phi = 0, \boldsymbol{\beta}\right)}{aver_{\Theta}\mathcal{L}_I\left(\mathbf{P}_{robust}; \lambda = 1, \phi = 0, \boldsymbol{\beta}\right)} \right) \times 100\%,$$

$$\text{Robustness payoff} = \left(1 - \frac{aver_{\Omega}\mathcal{L}_I\left(\mathbf{P}_{robust}; \lambda, \phi, \boldsymbol{\beta}\right)}{aver_{\Omega}\mathcal{L}_I\left(\mathbf{P}_{non-robust}; \lambda, \phi, \boldsymbol{\beta}\right)} \right) \times 100\%$$

where $\mathbf{P}_{non-robust}$ is the design constructed on the assumption that the fitted model is correct ($\phi = 0, \lambda = 1$) and $\mathbf{P}_{robust}$ is the robust design constructed by minimizing (4.16). The cost of the robustness of our designs is the lost efficiency. The payoff is the percentage reduction in loss due to the use of a robust design as opposed to a non-robust design which assumes the fitted model to be exactly correct. The efficiencies for the four designs are 100%, 98.5%, 98.3% and 96.9% while the payoffs are 0%, 0.21%, 11.5% and 35.3%, respectively. Thus the cost of robustness increases with the increasing width of the overdispersion parameter further away from zero which corresponds to no overdispersion. However, robustness is inexpensive here, the highest cost being 3.1% at the widest range $\Phi = [0, .5]$ and a corresponding payoff of 35.3%.

Now we fix the overdispersion parameter ϕ at zero and seek designs for four different range Λ of link parameter values with the range of model parameters kept at $\Theta = [.5, 1.5] \times [2.5, 3.5]$ as above. For all ranges of the link parameter considered the number of support points is two or three. These designs are presented in Figure 19. The robust designs protecting against misspecification are not characterized by a spreading of the support points over the design space as is the case of robust designs protecting against overdispersion. Rather these designs have fewer support points. The robustness is provided by redistribution and shifting of the location of these few design sites. The efficiencies of these robust designs are 99.4%, 94.9%, 95.9% and 97.6% for the range of link parameters $[.75, 1]$, $[.25, 1]$, $[0, 1]$, and $[0, 2]$ with payoffs

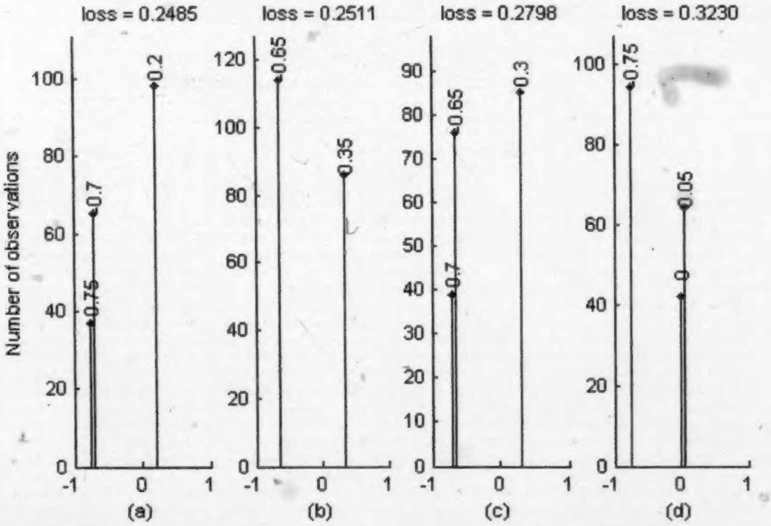

Figure 19: Minave designs when (a) $\Lambda = [.75, 1]$, (b) $\Lambda = [.25, 1]$, (c) $\Lambda = [0, 1]$, and (d) $\Lambda = [0, 2]$ with $\Theta = [.5, 1.5] \times [2.5, 3.5]$ and $\Phi = \{0\}$.

of .48%, 4.3%, 3.97% and 5.47%.

4.5 Application: Calandra Granaria Data

We return to Busvine's data on the mortality of grain beetle (*Calandra Granaria)* presented in Table 8. Using iterated reweighted least squares as discussed in Williams (1982), we fitted the overdispersion model (4.2). The parameter estimates are $\hat{\phi} = 0.946$, $\hat{\beta}_0 = -3.835$ and $\hat{\beta}_1 = 15.754$. Using the range $\Theta = [-2, -4] \times [8, 16]$ of parameter values, $\Phi = \{0\}$ and $\Lambda = \{1\}$ we construct the design minimizing (4.16). This is the design which assumes that there is no overdispersion nor link misspecification but still offers some robustness by incorporating a range of model parameter values. In Table 10 we labelled this the Regular Optimal Design. The model-robust design that protects against overdispersion as well as link misspecification is constructed by taking $\Theta = [-2, -4] \times [8, 16]$, $\Phi = [.75, 1]$ and $\Lambda = [0, 1]$. The regular optimal design is a 4-point design with support at dose levels .033, .167, .362 and .391. The model-robust design has support at all 10 dose levels in the design space. For comparison, we present two other designs in Table 10. The original design in the data for which there is no rationale given and the design with equal number of observation at all dose levels which we labeled "naive" design, being the design that a naive experimenter would consider. The results indicated that both the original design and the "naive" design are preferable to the regular optimal design since they have good robustness payoff with small efficiency loss. The robustness payoff is 63.7%, 63.9% and 66.8% for the original design, the "naive" design and the model-robust design, respectively. The model-robust design is superior of all four designs, it has the highest payoff at

Table 10: Comparing Designs for the Toxicity of Ethylene Oxide on Calandra Granaria.

Dose Level	Number of observations (Total: 290)			
	Original Design	"Naive" Design	Model-Robust Design	Regular Optimal Design
.033	24	29	62	5
.167	31	29	22	132
.199	30	29	19	–
.225	31	29	18	–
.260	27	29	18	–
.314	26	29	23	–
.322	30	29	24	–
.362	31	29	30	65
.391	30	29	37	88
.394	30	29	37	–
Lost Efficiency	6.5%	6.9%	4.8%	0%
Robustness Payoff	63.7%	63.9%	66.8%	0%

the least efficiency loss.

4.6 Concluding Remarks

The design criteria proposed in this work provide viable options to classical optimal design when there is the possibility of model misspecification. The numerical examples presented show that our designs offer some robustness against misspecification of the link function as well as overdispersion in binomial models. Adewale and Wiens (2005) reported increasing numbers of distinct observation sites with increasing model uncertainty in the linear predictor. They described the design that protects against uncertainty in the assumed form of the linear predictor as taking observations in clusters around the sites that would have been the design points for a classical design. Our findings in this work shows that the same is true for designs protecting against overdispersion. Generally, we found that the usual recommendation from classically optimal design would be inadequate in the presence of link misspecification and overdispersion. Our approach recommends a well spread out design over the design space would protect against overdispersion. However, there is a marked difference between the characteristics of designs protecting against overdispersion and those protecting against misspecification of link functions despite the fact that the same approach has been used for their construction. The recommendation from our procedure for robust design against misspecified link is a relocation of the few distinct design points from the design assuming correct model specification and possibly some redistribution of the number of observations at each of these sites.

While there have been many proposals for modelling extra-binomial variation, the approach we adopted is very attractive in that it accommodates both the random variation and the fixed effects on the same logistic scale - somewhat akin to a generalized linear mixed modelling approach. Another viable alternative that we consider is to specify the distribution of the true model as

$$Y_i|v_i \sim \frac{1}{n_i} binomial\left(n_i, \mu\left(\eta_i\right)\right), \ \eta_i = \mathbf{z}^T\left(\mathbf{x}_i\right)\boldsymbol{\beta}_0 + f\left(\mathbf{x}_i\right)$$

$$E\left(v_i\right) = \mu\left(\eta_i\right) \text{ and } var\left(v_i\right) = \phi^2\mu\left(\eta_i\right)\left(1-\mu\left(\eta_i\right)\right).$$

The mean response from this specification is exactly the approximate true mean response (4.3) used in this work but the variance is

$$var\left(Y_i\right) = \frac{1}{n_i} w\left(\mathbf{z}^T\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\left\{1 + \left(n_i - 1\right)\phi^2\right\}$$

which is slightly different from (4.4) given by the specification used in this work. Another way to obtain the same true mean response and variance prescribed by this alternative specification is to use a correlated binomial model:

$$\begin{aligned} Y_i &= \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}, \text{ with} \\ Y_{ij} &\sim \mathit{Bernoulli}\left(\mu\left(\eta_i\right)\right) \text{ and} \\ cov\left(Y_{ij}, Y_{ik}\right) &= \phi\mu\left(\eta_i\right)\left(1-\mu\left(\eta_i\right)\right), \; j \neq k. \end{aligned}$$

An advantage that comes with the correlated binomial model is that it can accommodate underdispersion by allowing negative correlation between the Bernoulli trials. Underdispersion, however, is a rare phenomenon in practice compared to overdispersion. Designs protecting against this form of alternative specification when $\phi > 0$ behave the same as our specification - a spread of the design points of the design space. Interestingly, when $\phi < 0$ under this specification, the resulting optimal design is similar to the classical optimal design recommendation - choosing a few distinct points to observe the response.

We note that the payoff for the robust design protecting against link misspecification may not always be impressive, as in the case of the example presented in this article. In the context of data analysis where this approach of treating a misspecified link as a systematic misspecification of the linear predictor was first proposed, Pregibon (1980) indicates "that the method is likely to be most useful for determining if a reasonable fit can be improved, rather than for the somewhat more optimistic goal of correcting a hopeless situation." Similarly, in a design context, the approach is meant to afford the experimenter some robustness when there is slight departure from the logit link. Further improvements might be obtained by using a second order Taylor's expansion

$$h(\mathbf{x};\lambda) = \left.\frac{\partial g}{\partial \lambda}\right|_{\lambda=1} (\lambda - 1) + \left.\frac{\partial^2 g}{\partial \lambda^2}\right|_{\lambda=1} (\lambda - 1)^2 + o\left([\lambda - 1]^2\right),$$

instead of (4.7).

While we have adopted the averaging of the loss function over uniformly scattered points from the parameter space in dealing with dependency of design criterion on unknown parameter, other possibilities remain. A proper Bayesian paradigm can be used which requires the specification of prior distribution on the parameters rather than the specification of a range of parameter values. It is our opinion that it might be easier to elicit information about plausible values of parameters from the experimenter than information leading to prior distribution assumptions. Another approach when an initial range of parameter values are available is the minimax approach. That is, an approach that seeks the design corresponding to the least loss for the worst possible parameter values in the specified range. Our experience is that the designs we presented here behave very similar to the minimax design (see Adewale and Wiens, 2006).

Acknowledgements

The author is grateful for guidance from his Ph.D. supervisor Prof. D.P. Wiens. The work was funded by the Natural Sciences and Engineering Research Council of Canada through a postgraduate scholarship to the author.

References

Abdelbasit, K.M. and Plackett, R.L. (1983), "Experimental Design for Binary Data," *Journal of the American Statistical Association*, 78, 90-98.

Adewale, A.J. and Wiens, D.P. (2006), "New Criteria for Robust Integer-Valued Designs in Linear Models," *Computational Statistics and Data Analysis*, in press.

Adewale, A.J. and Wiens, D.P. (2005), "Robust Designs for Misspecified Logistic Models," University of Alberta Statistics Centre Technical Report 05-08.

Atkinson, A.C. and Haines, L.M. (1996), "Designs for Nonlinear and Generalized Linear Models," in: *Handbook of Statistics*, ed. S. Ghosh and C.R. Rao, (13) pp. 437-475.

Biedermann, S., Dette, H. and Pepelyshev, A. (2004), "Some Robust Design Strategies for Percentile Estimation in Binary Response Models," Universität Dortmund, Technical Report 2004-19.

Bohachevsky, I.O., Johnson, M.E., and Stein, M.L. (1986), "Generalized Simulated Annealing for Function Optimization," *Technometrics*, 28, 209-217.

Box, G.E.P., and Draper, N.R. (1959), "A basis for the selection of a response surface design," *Journal of the American Statistical Association*, 54, 622-654.

Busvine, J.R. (1938), "The Toxicity of Ethylene Oxide to Calandra Oryzae, C. Granaria, Tribolium Castaneum, and Cimex Lectularius," *Annals of Applied Biology*, 25, 605-632.

Burridge, J. and Sebastiani, P. (1994), "D-optimal Designs for Generalized Linear

Models with Variance Proportional to the Square of the Mean," *Biometrika*, 81, 295-304.

Chaloner, K. and Larntz, K. (1989), "Optimal Bayesian Design Applied to Logistic Regression Experiments," *Journal of Statistical Inference and Planning*, 21, 191-208.

Chaudhuri, P. and Mykland, P. (1993), "Nonlinear experiments: Optimal Design and Inference Based on Likelihood," *Journal of the American Statistical Association*, 88, 538-546.

Chernoff, H. (1953), "Locally Optimal Designs for Estimating Parameters," *Annals of Mathematical Statistics*, 24, 586-602.

Crowder, M.J. (1978), "Beta-binomial Anova for Proportions," *Applied Statistics*, 27, 34-37.

Dean, C.B. (1992), "Testing for Overdispersion in Poisson and Binomial Regression Models," *Journal of the American Statistical Association*, 87, 451-457.

Dette, H. and Wong, W.K. (1996), "Optimal Bayesian Designs for Models with Partially Specified Heteroscedastic structure," *Annals of Statistics*, 24(5), 2108-2127.

Fang, K.-T. and Wang, Y. (1994), *Number-Theoretic Methods in Statistics*, Chapman & Hall, London.

Fang, Z., and Wiens, D.P. (2000), "Integer-valued, Minimax Robust Designs for Estimation and Extrapolation in Heteroscedastic, Approximately Linear Models," *Journal of the American Statistical Association*, 95, 807-818.

Ford, I, Titterington, D.M. and Kitsos, C.P. (1989), "Recent Advances in Nonlinear Experimental Design," *Technometrics*, 31, 49-60.

Ford, I. Torsney B. and Wu, C.F.J. (1992), "The Use of a Canonical Form in the Construction of Locally Optimal Designs for Nonlinear Problems," *Journal of the Royal Statistical Society* Series B, 54, 569-583.

Huber, P. (1975), "Robustness and Designs," in: *A Survey of Statistical Design and Linear Models*, ed. J.N. Srivastava, North Holland, Amsterdam, pp. 287-303.

King, J. and Wong, W.K. (2000), "Minimax D-Optimal Designs for Logistic Model," Biometrics, 56, 1263-1267.

Li, K.C., and Notz, W. (1982), "Robust Designs for Nearly Linear Regression," *Journal of Statistical Planning and Inference*, 6, 135-151.

Marcus, M.B., and Sacks, J. (1976), "Robust Designs for Regression Problems," in: *Statistical Theory and Related Topics II*, ed. S.S. Gupta and D.S. Moore, New York: Academic Press, pp. 245-268.

Minkin, S. (1987), "Optimal Design for Binary Data," *Journal of the American Statistical Association*, 82, 1098-1103.

Minkin, S. (1993), "Experimental Design for Clonogenic Assays in Chemotherapy," *Journal of the American Statistical Association*, 88, 410-420.

Pierce, D.A. and Sands, B.R. (1975), "Extra-Bernoulli Variation in Binary Data," Technical Report 46, Department of Statistics, Oregon State University.

Ponce de Leon, A. M. and Atkinson, A.C. (1993), "Designing Optimal Experiments for the Choice of Link Function for a Binary Data Model," in: *Model- Oriented Data Analysis*, ed. W. G. Muller, H. P. Wynn and A. A. Zhigljavsky, Heidelberg: Physica-Verlag, 25-36.

Pregibon, D. (1980), "Goodness of Link Selection in Generalized Linear Models," *Applied Statistics*, 29, 15 - 24.

Sitter, R.R. (1992), "Robust Designs for Binary Data," *Biometrics*, 48, 1145-1155.

Wiens, D.P. (1990), "Robust, Minimax Designs for Multiple Linear Regression," *Linear Algebra and Its Applications, Second Special Issue on Linear Algebra and Statistics*; 127, 327-340.

Wiens, D.P. (1992), "Minimax Designs for Approximately Linear Regression," *Journal of Statistical Planning and Inference*, 31, 353-371.

Wiens, D.P. (1998), "Minimax Robust Designs and Weights for Approximately Specified Regression Models with Heteroscedastic Errors," *Journal of the American Statistical Association*, 93, 1440-1450.

Williams, D.A. (1982), "Extra-binomial Variation in Logistic Linear Models," *Applied Statistics*, 31(2), 144-148.

CHAPTER V

CONCLUSIONS AND APPLICATIONS TO POISSON DATA

5.1 *Model-dependency of Optimal Designs*

This chapter discusses the results obtained for logistic models in Chapters 3 and 4 in the broader context of generalized linear models. For illustration of how the results apply to generalized linear models we again put the results in the framework of Poisson models. An Ames Salmonella reverse mutagenicity assay is used to illustrate various designs using the theories proposed in the previous chapters. The Poisson model is used for modelling responses involving counts. Designing for Poisson models has received only scant attention compare to the problem of designing for logistic regression models. Minkin (1993) and Behnken and Watts (1972) are among the few authors that have investigated designs for Poisson models.

We consider the data set compiled by Margolin et al. (1981) from an Ames Salmonella reverse mutagenicity assay. The number of revertant colonies was observed at six dose levels of quinoline. Three replicate plates were used at each dose level. The data are presented in Table 11.

Table 11: Number of revertant colonies of salmonella (Y_i).

	Y_i					
Obs.	x_i^*=0	10	33	100	333	1000
1	15	16	16	27	33	20
2	21	18	26	41	38	27
3	29	21	33	60	41	42

*Dose of quinoline (μmg/*plate*)

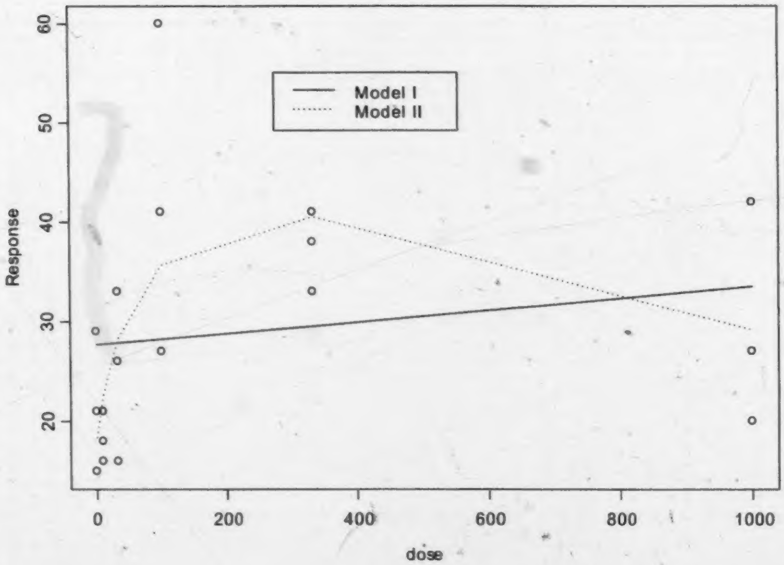

Figure 20: Number of revertant colonies of salmonella against dose level with Model *I* and *II* superimposed

A fit of the Poisson models with mean response

$$\begin{aligned} Model\ I &: \mu_1(x_i) = \exp\{\beta_0 + \beta_1 x_i\} \\ Model\ II &: \mu_2(x_i) = \exp\{\beta_0 + \beta_1 x_i + \beta_2 \log(x_i + 10)\} \end{aligned}$$

produce the following estimates: $\hat{\beta}_0 = 3.322$, $\hat{\beta}_1 = 0.0002$ (deviance = 75.81, df = 16) and $\hat{\beta}_0 = 2.173$, $\hat{\beta}_1 = -0.001$, $\hat{\beta}_2 = 0.320$ (deviance=43.72, df = 15), respectively. The deviance and corresponding degrees of freedom for each of the two fitted models indicate lack of fit. A plot of the data with the fitted models superimposed is presented in Figure 20. Figure 20 shows that the model with the term $\log(x_i + 10)$ in the linear predictor is an improved fit over the first model. However, the data presented variability that exceeds that prescribed by the Poisson distribution. There are many proposals in the literature for modelling overdispersed data. In order to accommodate overdispersion in Margolin's data, Lawless (1987) fitted the negative binomial model $Y_i \sim NB(\mu_i, k)$ with mean response

$$Model\ III : \mu(x_i) = \exp\{\beta_0 + \beta_1 x_i + \beta_2 \log(x_i + 10)\}$$

and variance $\mu(x_i) + k^{-1}\mu^2(x_i)$. The estimates obtained using the weighted least squares - method of moments are $\hat{\beta}_0 = 2.203$, $\hat{\beta}_1 = -0.001$, $\hat{\beta}_2 = 0.311$. A likelihood ratio test rejects the hypothesis $H : k^{-1} = 0$ and thus provides evidence of overdispersion in the data.

Consider an experimenter who intends to design an experiment to observe the number of revertant colonies of salmonella. For the sake of illustration, suppose subject matter knowledge restricts plausible dose levels to the six distinct dose levels

Table 12: Optimal Designs Minimizing Average Variance of Predictions.

Dose	0	10	33	100	333	1000
	Number of Observations					
Model I	12	–	–	–	–	6
Model II	5	–	–	4	4	5
Model III	5	–	–	4	5	4

presented in Table 11 and that the experimenter knows the parameters corresponding to models I, II and III. The designs presented in Table 12 are the designs minimizing the average variance of predictions over the six dose levels for each model. The design corresponding to Model I put two-third of the total observations at the control ($x = 0$) and the other one-third at the highest possible dose level ($x = 1000$). There is little difference between the designs corresponding to Model II and Model III. We note that each of these designs has been constructed with the salient assumption that the respective models are correct and that the experimenter has knowledge of the true parameters. In practice, there is always uncertainty or some form of misspecification in the assumed model. Thus the classical optimal designs taking the assumed model to be exactly correct could be inadequate since they do not protect against the possibility of alternative models. Simply put, the "optimal" design depends strongly on the assumed model.

5.2 Applications in Poisson Models

5.2.1 Poisson Models to be Fitted

Important applications of designs for Poisson regression models exist in toxicity studies, studies on the growth of algae over a time period in liquid samples and studies on penetration of materials by particles from a radioactive source. Minkin (1993) investigated design for Poisson regression models when the response is the number of colonies formed over different concentrations of an anticancer drug.

Suppose the experimental considers fitting a Poisson model with mean count

$$E(Y_i) = var(Y_i) = \mu_i = e^{\eta_i}, \text{ with } \eta_i = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta}, \tag{5.1}$$

where $\mathbf{z}(\mathbf{x}_i) = (1, \mathbf{x}_i)^T$ is the vector of regressors. In a dose-response study, for example, x_i would denote a dose level. The experimenter needs to choose the dose levels at which to observe the response from a set of dose levels $\mathcal{S} = \{x_i\}_{i=1}^N$ and the number of observations to be taken at each dose level. We assume that the set $\mathcal{S}$ is finite. Given the total number of observations n desired, the classical design problem is then to choose the number of observations $n_1, ..., n_N$ in some optimal manner with respect to the model (5.1).

Knowing that in practice the model (5.1) is at best an approximation to the true unknown model, in what follows we consider three forms of misspecifications and construct robust designs that protect against these forms of misspecification.

5.2.2 Departures from the Assumptions and Design Criteria for Protection Against Departures

Model misspecifications have implications for both the asymptotic bias and asymptotic covariance matrix of the model parameter estimates. In the spirit of the pioneering work of Box and Draper (1959) on model-robust designs, we employ the mean squared error of predictions averaged over the design space as our design criterion. We use the normalized average mean squared error (AMSE) I of the response prediction

$\mu(\hat{\eta}_i)$, with $\hat{\eta}_i = \mathbf{z}^T(\mathbf{x}_i)\hat{\boldsymbol{\beta}}$, as the design criterion,

$$I \triangleq \frac{n}{N}\sum_{i=1}^{N} E\left[\left\{E\left[\hat{Y}|\mathbf{x}_i\right] - E\left[Y|\mathbf{x}_i\right]\right\}^2\right]. \tag{5.2}$$

5.2.2.1 Linear Predictor Misspecification in Poisson Models

Now, suppose the true model has the mean response

$$E_T(Y_i) = \mu_{T,i} = e^{\mathbf{z}^T(\mathbf{x}_i)\beta + f(\mathbf{x}_i)},$$

with $\mathbf{z}(\mathbf{x}_i) = (1, \mathbf{x}_i)^T$ and that f belongs to a contamination neighbourhood $\mathcal{F}$,

$$\mathcal{F} = \left\{\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}(\mathbf{x}_i) f(\mathbf{x}_i) = 0,\ \frac{1}{N}\sum_{i=1}^{N} f^2(\mathbf{x}_i) \le \tau^2, \text{ with } \tau^2 = O\left(n^{-1}\right)\right\}. \tag{5.3}$$

Thus the model (5.1) has a misspecified linear predictor. In Chapter 2 we show that the asymptotic mean squared error depends on the unknown contamination f. The dependency of the design criterion on unknown contamination was resolved by averaging the loss over the contamination neighbourhood (5.3). The averaging is carried out using a procedure based on the singular value decomposition of the design matrix $\mathbf{Z}$ (see also Chapter 1 and Fang and Wiens, 2000). The average loss over the misspecification neighbourhood $\mathcal{F}$ is given by

$$\begin{aligned}\mathcal{L}_{I,ave}(\mathbf{P},\rho) &\triangleq \frac{1}{N} tr\left[\left(\mathbf{U}^T\mathbf{PWU}\right)^{-1}\left(\mathbf{U}^T\mathbf{W}^2\mathbf{U}\right)\right] \\ &\quad + \frac{\rho}{N-p+2} tr\left[\mathbf{W}\left(\mathbf{R}-\mathbf{I}_N\right)\left(\mathbf{R}^T-\mathbf{I}_N\right)\mathbf{W}\right],\end{aligned} \tag{5.4}$$

where $\rho = n\tau^2$, $\mathbf{U}$ is obtained from the singular value decomposition of the design matrix $\mathbf{Z} = \mathbf{U}_{N\times p}\boldsymbol{\Lambda}_{p\times p}\mathbf{V}_{p\times p}^T$ and $\mathbf{R}_{N\times N} = \mathbf{U}\left(\mathbf{U}^T\mathbf{PWU}\right)^{-1}\mathbf{U}^T\mathbf{PW}$.

Thus for any given value of the misspecification parameter, ρ, designs protecting against linear predictor misspecification are obtained by choosing the matrix of allocations $\mathbf{P}$ such that (5.4) is minimized.

5.2.2.2 Overdispersion in Poisson Models

Minkin (1993) consider overdispersion when constructing design for Poisson data. He accommodates overdispersion using a multiplicative random effect. That is, by taking the true model as $Y_i|v \sim \text{Poisson}(v\mu_i)$, where the random variable v has mean 1 and variance ϕ^2. Thus the unconditional distribution of Y_i has the mean response is μ_i and variance $\mu_i + \phi^2\mu_i^2$. In this work we rather accommodate overdispersion by an additive random contamination to the linear predictor such that the true model has the conditional distribution $Y_i|v \sim \text{Poisson}(\mu_{T,i} = e^{\eta_i})$ with $\eta_i = \mathbf{z}^T(x_i)\boldsymbol{\beta}+\phi v$ where v is a random variable with mean 0 and variance 1. Thus, the true model has mean

$$E(Y_i) = \mu_{T,i} \approx \mu_i = e^{\mathbf{z}^T(t_i)\boldsymbol{\beta}}$$

and variance

$$var_T(Y_i) \approx \mu_i + \phi^2\mu_i^2.$$

The resulting mean response is therefore approximately equal to the mean response and variance of the multiplicative random effect approach. This approach was adopted just to keep misspecification to the linear predictor and thus unify our results for all kinds of misspecification considered. Similar approach is used to tend to overdispersion when designing for binomial models with overdispersion in Chapter 4. The average asymptotic mean squared error of predictions, which is the same as the average variance of predictions here, is given by

$$\mathcal{L}_I(\mathbf{P}) = \frac{1}{N}\left\{tr\left[\mathbf{WZH}_n^{-1}\tilde{\mathbf{H}}_n\mathbf{H}_n^{-1}\mathbf{Z}^T\mathbf{W}\right]\right\} \tag{5.5}$$

For the model with $E(Y_i) = \mu_{T,i} \approx \mu_i$ and $var_T(Y_i) \approx \mu_i + \phi^2\mu_i^2$, we have $w_{T,i} = w_i + \phi^2 w_i^2$ so that (5.5) simplifies to

$$\mathcal{L}_I(\mathbf{P}, \phi) = \frac{1}{N}\left\{tr\left[\mathbf{WZH}_n^{-1}\mathbf{Z}^T\mathbf{W}\right] + \phi^2 tr\left[\mathbf{WZH}_n^{-1}\mathbf{G}_n\mathbf{H}_n^{-1}\mathbf{Z}^T\mathbf{W}\right]\right\}, \tag{5.6}$$

where $\mathbf{G}_n = \mathbf{Z}^T\mathbf{WPWZ}$. Given a range Φ of plausible values for the overdispersion parameter ϕ, our design criterion is the average of $\mathcal{L}_I(\mathbf{P}, \phi)$ over Φ, that is,

$$\mathcal{L}_{I,ave}(\mathbf{P}) = aver_{\Phi}\mathcal{L}_I(\mathbf{P}, \phi). \tag{5.7}$$

5.2.2.3 Link Misspecification in Poisson Models

In designing for a Poisson model with possibly misspecified link function we employ the same framework as in Chapter 4 where the true link function and the fitted canonical link belong to a generalized family of link functions. Here consider the parameterized family of link functions defined by

$$g(\mu, \lambda) = \begin{cases} \mu^{\lambda}, & \lambda \neq 0 \\ \log\mu, & \lambda = 0 \end{cases}, \tag{5.8}$$

where λ is the link parameter. The log link corresponds to $\lambda = 0$. The strategy is to linearize this generalized link function by expanding it in a Taylor series about a fixed value $\lambda = 0$ and taking only the linear term. We then approximate the true linear predictor by

$$\eta_i = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta}_0 + f(\mathbf{x}_i, \lambda), \tag{5.9}$$

with the contamination function $f(\mathbf{x}; \lambda)$ given by

$$f(\mathbf{x}, \lambda) = -\left.\frac{\partial g(\mathbf{x}, \lambda)}{\partial \lambda}\right|_{\lambda=0} \lambda = -\lambda \log \mu. \tag{5.10}$$

This approach follows the work of Pregibon (1980) on goodness of link tests for generalized linear models.

Our criterion for designs protecting against link misspecification is given by

$$\mathcal{L}_I(\mathbf{P}; \lambda) = \frac{1}{N}\left\{tr\left[\mathbf{W}\mathbf{Z}\mathbf{H}_n^{-1}\tilde{\mathbf{H}}_n\mathbf{H}_n^{-1}\mathbf{Z}^T\mathbf{W}\right] + n\left\|\mathbf{W}\left(\mathbf{Z}\mathbf{H}_n^{-1}\mathbf{b} - \mathbf{f}\right)\right\|^2\right\} \tag{5.11}$$

for $\mathbf{f} = (-\lambda \log \mu_1, ..., -\lambda \log \mu_N)^T$ (see Corollary 3.1). Given a range Λ of plausible values of the link parameter λ we seek designs that minimize the average of $\mathcal{L}_I(\mathbf{P}; \lambda)$ over Λ, that is,

$$\mathcal{L}_{I,ave}(\mathbf{P}) = aver_{\Lambda}\mathcal{L}_I(\mathbf{P}; \lambda). \tag{5.12}$$

5.2.3 Summary of Implications of Model Misspecifications

The maximum likelihood estimates of the model parameters are no longer asymptotically unbiased when there is model misspecification. In the same vein, the asymptotic covariance matrix of the maximum likelihood estimates of the model parameters is no longer equal to the inverse of the Fisher's information matrix. This position is congruent with the work of White, H. (1982) and Fahrmeir, L. (1990) on maximum likelihood estimation of misspecified models.

The approximations for the mean response and variance of the true model has implications for the terms involved in the components of the average mean squared error loss function (5.2). When there is linear predictor misspecification (misspecification of fixed effects), the true mean response depends on the contamination function

$f(\cdot)$ through the linear predictor. Since the variance of the response is a function of the mean in generalized linear models, the covariance matrix of the maximum likelihood estimates of model parameters also depends on the unknown contamination function $f(\cdot)$ through the mean. When the true model is an overdispersed model (of the form assumed in this dissertation), the true mean response is approximately the same as the fitted mean response. The maximum likelihood estimates of model parameters remains asymptotically unbiased for small value of the overdispersion parameter. Asymptotically, the overdispersion will only affect the covariance matrix of the maximum likelihood estimates of model parameters. For the case of misspecified link function, the linearization of the generalized family of links has the implication that the linear predictor is contaminated with a function which is the derivative of the true link function evaluated at the link parameter value for the canonical link function. Thus the true mean response (and the variance function) depends on this contamination function.

Therefore the misspecification of the fixed effects and the link function affects the asymptotic bias and covariance matrix of the maximum likelihood estimates of model parameters since both kinds of misspecifications are invariably considered as systematic misspecification of the linear predictor via the contamination function $f(\cdot)$. The difference between these two forms of misspecifications, however, is that the contamination function emanating from fixed effects misspecification is an unknown member of the contamination neighbourhood $\mathcal{F}$ defined by (5.3) while the contamination function due to link misspecification is a known function of the fitted mean response. It is conceptually reasonable to assume that the problem of link misspecification and fixed effect misspecification are mutually exclusive. That is, the experimenter should seek protection against link misspecification only if there is assurance that the linear combination of regressors (the fixed effects) have been adequately taken care of.

5.3 Computational Issues

As in the general problem of design for nonlinear models, the design criteria in this dissertation also depend on unknown model parameters. In dealing with the parameter dependencies we have employed approaches such as assuming best guess values (see Chapter 3), assuming a prior distribution on the parameter and then optimize an expectation (with respect to the assumed prior distribution) of the design criterion (see Chapter 3) and assuming a range of plausible parameter values and then optimizing an average value of the criterion over the assumed range (see Chapter 4). The evaluation of expectation (when there is prior distribution of parameter) and averaging (when there is a range of plausible parameter values) is carried out numerically using quasi-Monte Carlo procedures (see Fang and Wang, 1994). The quasi-Monte Carlo method (or Number-Theoretic method) is based on the notion of choosing a set of points which are uniformly scattered in the s-dimensional unit cube $\mathcal{C}^s$. It should be noted that "uniformly scattered" set is not in the usual statistical sense of uniformly distributed points, rather it is in the sense that for any neighbourhood $\mathcal{D}_\Omega$ of the parameter range Ω the discrepancy defined by

$$discrepancy = \sup_{\mathcal{D}_\Omega \subseteq \Omega} \left| \frac{\#\{\mathcal{D}_\Omega\}}{\#\{\Omega\}} - \frac{vol\,(\mathcal{D}_\Omega)}{vol\,(\Omega)} \right|$$

is small, where $\#\{\cdot\}$ denotes the cardinality of $\{\cdot\}$ and $vol\,(\cdot)$ denotes its volume. There are many proposals in the number-theoretic method literature for generating these uniformly scattered set of points over various domains. See Fang and Wang (1994) for relevant theory and discussion on the applications of quasi-Monte Carlo method in statistics. The main attraction for their use in this work is that they overcome the inefficiency of the Monte Carlo method when the number of random sample being taken is small. Also, quasi-Monte Carlo method is somewhat deterministic (nonrandom) and thus guarantees reproducibility of designs even when the sample size of uniformly scattered set is small.

In addition to the dependency of the design criteria on unknown model parameters, the criteria developed in this dissertation also requires the specification of the design parameters. Designing for misspecified linear predictor using the minave approach developed in this dissertation requires the specification of the parameter ρ. The averaging over the contamination neighbourhood is handled analytically through a characterization of the neighbourhood which cast the average loss as an expected value of the loss function with respect to the uniform measure on a unit sphere and its interior. See chapter 3 for the details of this development. On the other hand, designing for generalized linear models with overdispersion and possible link misspecification using the minave approach requires the specification of a range of plausible parameter values. In this case, the averaging is done using the quasi-Monte Carlo method described earlier. It should be recalled that the design parameters are measures of the experimenter's belief on the departure of the true but unknown model from the model he intends to fit. Some guides for choosing these design parameters are presented in §5.5.

The minimization of the losses (5.4), (5.7) and (5.12) are nonlinear integer optimization problems. In this dissertation we have employed the simulated annealing algorithm for the resulting optimization problems. The simulated annealing algorithm had been used in optimal design problems by, among others, Haines (1987), Meyer and Nachtsheim (1988) and Fang and Wiens (2000). Even though the implementation of the algorithm varies from author to author and application to application, the general principle remains three simple steps: (i) specify an initial design, (ii) develop a basis of generation of new design by a small perturbation of the current design, and (iii) specify a probabilistic criterion to determine if a new design is accepted or rejected. The details of our implementation is described in Chapters 2, 3 and 4. All examples in this dissertation work are coded in matlab.

5.4 Application to Ames Salmonella Assay

The design in the Ames Salmonella assay presented in Table 11 has six dose levels and 3 replicates observations at each dose level. There is no rationale given for this design. In this section we consider designing for this assay when the interest is to take a total of 18 observations and the design space consists of 50 equally spaced points between 0 and 1000 μmg (including the points 0 and 1000).

Suppose the experimenter intends to fit the Poisson model with mean response

$$\mu = \exp\left(\eta_i\right), \tag{5.13}$$

with the linear predictor

$$\eta_i = \beta_0 + \beta_1 x_i + \beta_2 \log\left(x_i + 10\right) \tag{5.14}$$

and that he desires to have some protection against possible misspecification of this linear predictor. The ranges of plausible parameter values are taken to be $[2.0, 3.5]$, $[-.002, .001]$, and $[.30, .35]$ for β_0, β_1 and β_2, respectively. Note that these ranges of parameter values cover the estimates from the fit of Model II and Model III using the data in Table 11. The designs presented in Table 13 minimizes the average of (5.4) over uniformly scattered set of points from $\Theta = [2.0, 3.5] \times [-.002, .001] \times [.30, .35]$ for given values of the design parameter ρ. The design parameter $\rho = 0$ corresponds to the assumption that the model defined by (5.13) and (5.14) is exactly correct. We evaluate our designs using their efficiencies. The efficiency of a design represented by the allocation matrix $\mathbf{P}$ is given by:

$$\text{Efficiency} = \left(\frac{aver_{\beta\in\Theta}\mathcal{L}_{I,ave}\left(\mathbf{P}_{non-robust}, \rho = 0, \boldsymbol{\beta}\right)}{aver_{\beta\in\Theta}\mathcal{L}_{I}\left(\mathbf{P}; \rho = 0, \boldsymbol{\beta}\right)} \right) \times 100\%,$$

where the allocation matrix $\mathbf{P}_{non-robust}$ corresponds to the design constructed assuming that there is no misspecification which the experimenter hopes to protect against. The efficiency of the designs for $\rho = .05, .25$ and $.5$ relative to $\rho = 0$ are 98%, 95.8%, and 88.4%, respectively.

As with logistic models with misspecified linear predictors, the robust designs for Poisson models with misspecified linear predictor are designs with clusters of support points around the support points of a design that assumes the fitted model is exactly correct.

Now suppose the experimenter still intends to fit the Poisson model with the mean response (5.13) and the linear predictor (5.14) but seeks protection against overdispersion. The designs protecting against overdispersion for various ranges Φ

Table 13: Designs for a Poisson Model with Misspecified Linear Predictor.

ρ	Support points (Number of observations)	Loss	Efficiency
0	0 (2), 306.12 (4), 326.53 (5), 1000 (7)	219.15	100
.005	0 (2), 224.49 (1), 244.9 (1), 265.31 (1), 285.71 (1) 306.12 (1), 326.53 (1), 346.94 (1), 367.35 (1) 387.76 (1), 918.37 (1), 938.78 (1), 959.18 (1) 979.59 (2), 1000 (2)	335.42	98
.05	0 (2), 204.08 (1), 244.9 (1), 265.31 (1), 285.71 (1) 306.12 (1), 326.53 (1), 346.94 (1), 367.35 (1) 387.76 (1), 877.55 (1), 897.96 (1), 918.37 (1), 938.78 (1), 959.18 (1), 979.59 (1), 1000 (1)	1314.69	95.9
.5	0 (1), 224.49 (1), 265.31 (1), 285.71 (1) 306.12 (1), 326.53 (1), 346.94 (1), 367.35 (1) 387.76 (1), 408.16 (1), 857.14 (1), 877.55 (1), 897.96 (1), 918.37 (1), 938.78 (1), 959.18 (1), 979.59 (1), 1000 (1)	11035.78	88.4

These designs are computed with the range $\Theta = [2.0, 3.5] \times [-.002, .001] \times [.30, .35]$ assumed for the model parameters β_0, β_1 and β_2.

of the overdispersion parameter are presented in Table 14. These designs have few distinct support points and the number of support points is not increasing with the increasing departure of the overdispersion parameter from the value corresponding to no overdispersion. The corresponding designs protecting against link misspecification are presented in Table 15 for various ranges Λ of the link parameter.

The lost efficiency increases as the design parameter moves away from the value which implies that the assumed model is exactly correct.

5.5 Choosing Design Parameters using Efficiency

The design parameter ρ in the criterion (5.4) is a measure of the extent of linear predictor misspecification. In the Ames Salmonella assay application of §5.4, we present

Table 14: Designs for a Poisson Model with Protection Against Overdispersion.

Φ	Support points (Number of observations)	Loss	Efficiency
$\{0\}$	0 (2), 306.12 (4), 326.53 (5), 1000 (7)	223.63	100
$[.05, .1]$	0 (2), 285.71 (1), 306.12 (7), 1000 (8)	462.55	98.0
$[.25, .28]$	0 (1), 285.71 (8), 1000 (9)	2939.41	88.1
$[.5, 1.0]$	0 (1), 285.71 (8), 1000 (9)	23137.45	88

These designs are computed with the range $\Theta = [2.0, 3.5] \times [-.002, .001]$ $\times [.30, .35]$ assumed for the model parameters β_0, β_1 and β_2.

Table 15: Designs for a Poisson Model with Misspecified Link Function.

Λ	Support points (Number of observations)	Loss	Efficiency
$\{0\}$	0 (2), 306.12 (4), 326.53 (5), 1000 (7)	223.63	100
$[.5, .15]$	0 (2), 244.9 (5), 265.31 (3), 530.61 (1), 734.69 (1), 1000 (6)	18351.1	97.7
$[.25, .5]$	(7), 20.41 (9), 1000 (2)	725986.03	10.6
$[.5, 1.0]$	0 (8), 20.41 (8), 1000 (2)	6367114.79	10.5

These designs are computed with the range $\Theta = [2.0, 3.5] \times [-.002, .001]$ $\times [.30, .35]$ assumed for the model parameters β_0, β_1 and β_2.

designs corresponding to various values of ρ. In practice this parameter would be assigned a value that represent the belief of the experimenter about the departure of the true linear predictor from that which the experimenter intends to fit. In the same vein the parameters ϕ and λ in (5.6) and (5.11), respectively, are measures of the extent of the overdispersion and link misspecification. In the examples we assumed the experimenter knows the value of ρ in the case of designing for misspecified linear predictor and a range of plausible values for the overdispersion and link parameters. This knowledge can be based on prior data or just similar experiments reported in the

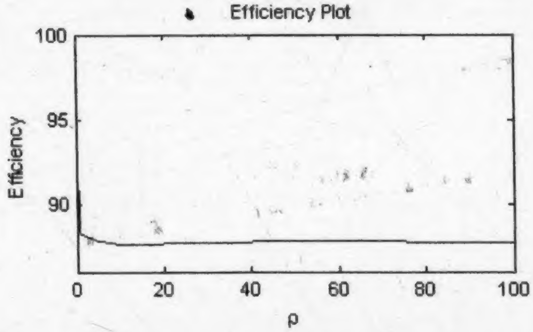

Figure 21: Efficiency of robust designs for a Poisson model with misspecified linear predictor over a grid of values of ρ in the interval $[0, 100]$

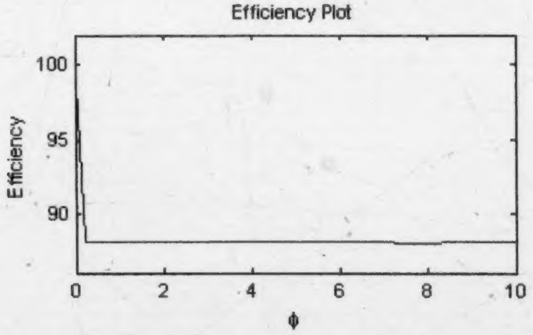

Figure 22: Efficiency of robust designs Poisson model with overdispersion

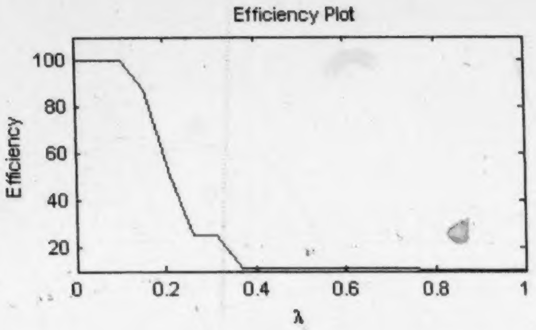

Figure 23: Efficiency of robust designs for a Poisson model with misspecified link function over a grid of values of λ in the interval $[0, 1]$

literature. Here we propose an alternative approach based on efficiency to choosing these design parameters (linear predictor misspecification parameter, ρ, overdispersion parameter, ϕ and link misspecification parameter, λ).

Given the finite design space $\mathcal{S}$, the number of observations desired, the contemplated model (to be fitted) and the range of model parameter values, the experimenter employs the criterion corresponding to the kind of departure he needs protection against to construct designs over a grid of values of the corresponding design parameter (ρ, ϕ, or λ). The experimenter can then choose a robust design corresponding to the level of efficiency desired while noting that robustness is at the expense of efficiency. The design corresponding to the assumption that the contemplated model is exactly correct is 100% efficient. He then construct designs for various values of relevant design parameter and evaluate the efficiencies of the designs corresponding to these design parameter values. The recommended value(s) of the design parameter is that which corresponds to the level of efficiency the experimenter is willing to sacrifice for robustness.

As an illustrative example, using the relevant design criterion under the setting of the Ames Salmonella assay of §5.4 we plot the efficiencies of the designs corresponding to grids of design parameters ρ (Figure 21), ϕ (Figure 22), and λ (Figure 23). For all the three cases the efficiency plot tails-off after some value of the design parameter. This means that the designs stay constant after some value of the design parameter. The least efficiency from the robust design for misspecified linear predictor is 87.6%, 88.1% for the robust design for overdispersion and 10.7% for the robust design for link misspecification. The procedure for the link misspecification case is only efficient at values of the link parameter λ near the canonical link value $\lambda = 0$ intended to be fitted. Thus if the experimenter thinks that the canonical link may be grossly inadequate designs using a more plausible link function should be sought. This observation corroborates the position of Pregibon (1980) that the link linearization procedure is not meant for an optimistic goal of correcting a seemingly hopeless situation.

Appendix - Theorem 3.1 stated for Generalized Linear Models

Preliminaries:

Consider a finite design space $\mathcal{S} = \{\mathbf{x}_i\}_{i=1}^N$ of possible design points from which an experimenter is interested in choosing n, not necessarily distinct, points at which to observe a response Y. The experimenter makes $n_i \geq 0$ observations at x_i such that $\sum_{i=1}^N n_i = n$. The design problem is to choose $n_1, ..., n_N$ in an optimal manner. The objective then is to choose a probability distribution $\{p_i\}_{i=1}^N$, with $p_i = n_i/n$, on the design space $\mathcal{S}$. The log-likelihood l, according to the possibly misspecified model is

$$\begin{aligned} l(\boldsymbol{\beta}) &= \sum_{i=1}^N l_i = \sum_{i=1}^N \log f(Y_i; \theta_i, \phi) \\ &= \sum_{i=1}^N \{[Y_i\theta_i - b(\theta_i)]/a(\phi) + c(Y_i, \phi)\}. \end{aligned}$$

We employed the chain rule to obtain the first derivative of the log-likelihood l,

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^N \frac{\partial l_i}{\partial \boldsymbol{\beta}} = \sum_{i=1}^N \frac{\partial l_i}{\partial \theta_i}\frac{\partial \theta_i}{d\mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \boldsymbol{\beta}},$$

where $\frac{\partial l_i}{\partial \theta_i} = [Y_i - b'(\theta_i)]/a(\phi)$. Using $\mu_i = b'(\theta_i)$ and $var(Y_i) = a(\phi)b''(\theta_i)$ we have

$$\begin{aligned} \frac{\partial l_i}{\partial \theta_i} &= [Y_i - \mu_i]/a(\phi), \\ \frac{\partial \mu_i}{d\theta_i} &= b''(\theta_i) = var(Y_i)/a(\phi), \text{ and} \\ \frac{\partial \eta_i}{\partial \boldsymbol{\beta}} &= \mathbf{z}(\mathbf{x}_i) \text{ since } \eta_i = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\beta}. \end{aligned}$$

Thus

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^N \frac{(Y_i - \mu_i)}{var(Y_i)}\frac{\partial \mu_i}{\partial \eta_i}\mathbf{z}(\mathbf{x}_i)$$

and

$$\frac{\partial^2 l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} = \sum_{i=1}^{N}\left[\frac{\partial\mu_i/\partial\eta_i}{var(Y_i)}\mathbf{z}(\mathbf{x}_i)\frac{\partial\{Y_i-\mu_i\}}{\partial\boldsymbol{\beta}^T} + (Y_i-\mu_i)\,\mathbf{z}(\mathbf{x}_i)\frac{\partial}{\partial\boldsymbol{\beta}^T}\left\{\frac{\partial\mu_i/\partial\eta_i}{var(Y_i)}\right\}\right].$$

Claim 1:

$$\frac{\partial\{Y_i-\mu_i\}}{\partial\boldsymbol{\beta}^T} = -\frac{\partial\mu_i}{\partial\eta_i}\mathbf{z}^T(\mathbf{x}_i).$$

Claim 2: For a canonical model (which is the fitted model)

$$\frac{\partial}{\partial\boldsymbol{\beta}^T}\left\{\frac{\partial\mu_i/\partial\eta_i}{var(Y_i)}\right\} = \mathbf{0}.$$

Thus, -1 times the second derivative of second derivative of the possibly misspecified canonical model is

$$-\frac{\partial^2 l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} = \sum_{i=1}^{N}\frac{\partial\mu_i/\partial\eta_i}{a(\phi)}\mathbf{z}(\mathbf{x}_i)\,\mathbf{z}^T(\mathbf{x}_i).$$

The preceding arguments cumulate into Theorem 5.1 which gives the expressions for the asymptotic bias and asymptotic covariance of the maximum likelihood estimate of the model parameters.

Theorem 5.1 *Let* $\boldsymbol{\gamma}$ *and* $\boldsymbol{\gamma}_T$ *be the* $N \times 1$ *vectors with elements* μ_i *and* $\mu_{T,i}$ *respectively, where* μ_i *is the fitted mean response and* $\mu_{T,i}$ *is the true mean response and let* $\mathbf{Z}$ *be the* $N \times p$ *matrix with rows* $\mathbf{z}^T(\mathbf{x}_i)$*. Let* $var_T(Y_i)$ *be the true variance. Define*

$$w_i = \frac{\partial\mu_i/\partial\eta_i}{a(\phi)}, \text{ and } w_{T,i} = \frac{var_T(Y_i)}{a^2(\phi)}. \tag{5.15}$$

Let $\mathbf{P}$, $\mathbf{W}$, $\mathbf{W}_b$ *and* $\mathbf{W}_T$ *be the* $N \times N$ *diagonal matrices with diagonal elements* n_i/n, w_i, $w_i\partial\eta_i/\partial\mu_i$ *and* $w_{T,i}$ *respectively. Finally, define*

$$\begin{aligned}
\mathbf{b} &= \mathbf{Z}^T\mathbf{P}\mathbf{W}_b(\boldsymbol{\gamma}_T-\boldsymbol{\gamma}),\\
\mathbf{H}_n &= \mathbf{Z}^T\mathbf{P}\mathbf{W}\mathbf{Z},\\
\tilde{\mathbf{H}}_n &= \mathbf{Z}^T\mathbf{P}\mathbf{W}_T\mathbf{Z}.
\end{aligned}$$

The asymptotic bias and asymptotic covariance matrix of the maximum likelihood estimate $\hat{\boldsymbol{\beta}}$ *of the model parameter vector* $\boldsymbol{\beta}$ *from the misspecified model are*

$$\begin{aligned} bias\left(\hat{\boldsymbol{\beta}}\right) &= E\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) = \mathbf{H}_n^{-1}\mathbf{b} + o\left(n^{-1/2}\right), \\ cov\left(\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)\right) &= \mathbf{H}_n^{-1}\tilde{\mathbf{H}}_n\mathbf{H}_n^{-1} + o\left(1\right), \end{aligned}$$

respectively.

Proof: This is a generalization of Theorem 3.1 to generalized linear models. See Chapter 3 for a proof specific to logistic model.

References

Adewale, A.J. and Wiens, D.P. (2005), "Robust Designs for Misspecified Logistic Models," University of Alberta Statistics Centre Technical Report 05-08.

Behnken, D.W. and Watts, D.G. (1972), "Bayesian Estimation and Design of Experiments for Growth Rates when Sampling from the Poisson Distribution," *Biometrics*, 28, 999-1009.

Fahrmeir, L. (1990), "Maximum Likelihood Estimation in Misspecified Generalized Linear Models," *Statistics*, 21 (4), 487-502.

Fang, K.-T. and Wang, Y. (1994), *Number-Theoretic Methods in Statistics*, Chapman & Hall, London.

Fang, Z., and Wiens, D.P. (2000), "Integer-Valued, Minimax Robust Designs for Estimation and Extrapolation in Heteroscedastic, Approximately Linear Models," *Journal of the American Statistical Association*, 95, 807-818.

Haines, L.M. (1987), "The Application of the Annealing Algorithm to the Construction of Exact Optimal Designs for Linear-Regression Models," *Technometrics*, 29, 439-447.

Lawless, J.F. (1987), "Negative Binomial and Mixed Poisson Regression," *The Canadian Journal of Statistics*, 15, 209-225.

Margolin, B.H., Kaplan, N., and Zeiger, E. (1981), "Statistical Analysis of the Ames Salmonella/Microsome Test," Proc. Natl. Acad. Sci. USA, 76, 3779-3783.

Meyer, R.K., and Nachtsheim, C.J. (1988), "Constructing Exact D-Optimal Experimental Designs by Simulated Annealing," *American Journal of Mathematical and Management Sciences*, 3 & 4, 329-359.

Minkin, S. (1993), "Experimental Design for Clonogenic Assays in Chemotherapy," *Journal of the American Statistical Association*, 88, 410-420.

Pregibon, D. (1980), "Goodness of Link Selection in Generalized Linear Models," *Applied Statistics*, 29, 15 - 24.

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-25.